Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Direction of Shift in the Judgment of Single Stimuli: A. PARDUCCI 169

Time Uncertainty in Simple Reaction Time: E. T. KLEMMER 179

Changes in Response Strength with Changes in the Amount of Reinforcement: R. H. DUFORT AND G. A. KIMBLE 185

Conditioned Stimulus Intensity and Temporal Factors in Spaced-Trial Classical Conditioning: G. W. BARNES 192

Learning and Retaining a Rate of Movement with the Aid of Kinesthetic and Verbal Cues: R. S. LINCOLN 199

Spontaneous Recovery and Statistical Learning Theory: L. E. HOMME 205

"Backward" Learning in Paired Associates: B. B. MURDOCK, JR. 213

Verbal Learning and Retention as a Function of the Number of Competing Associations: R. S. BERCHDT 216

Importance of Object Recognition in Size Constancy: R. C. BOLLES AND D. E. BAILEY 222

Effect of Delay of Knowledge of Results on Learning a Motor Task: J. GREENSPOON AND S. FOREMAN 226

American Psychological Association

Vol. 51 No. 3 March 1956

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1956 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright © 1956 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 51, NO. 3 MARCH, 1956

DIRECTION OF SHIFT IN THE JUDGMENT OF SINGLE STIMULI [1]

ALLEN PARDUCCI [2]

University of Oregon

When a change is made in the *range* of the stimuli being presented for judgment, *Ss* typically change their adaptation level and general judgment scale for the stimulus dimension they are judging. Introduction of new stimuli which are more extreme, larger, heavier, or longer than any of the stimuli already presented produces a shift in the judgments applied to the old stimuli. These rejudgments tend to employ less extreme categories, the old stimuli being judged smaller, lighter, or shorter than they had been before the extension of the stimulus range (e.g., **2, 6, 7**).

Principles describing these shifts in judgment have been elaborated by Helson (**2**), Johnson (**4**), and Volkmann (**9**). Their formulations suggest also that if the range of the stimuli were restricted instead of extended, e.g., if the largest, heaviest, or longest stimuli were dropped out of the series being presented for judgment, the judgments of the remaining stimuli would shift in the opposite direction. Thus, *Ss* would tend to apply the extreme judgment categories to the most extreme of the remaining stimuli.

Although there appear to have been no experimental studies of shifts in judgment following *restriction* of the stimulus range, this would seem to be a fruitful area for research because of the special opportunities for varying the *discriminability* of the stimulus shifts. If the stimuli which were dropped out had appeared only infrequently during the preshift series, their absence might not be expected to produce an immediate shift in *Ss'* judgments. The experiments to be reported in this paper attempted to study the effects upon judgment of different stimulus shifts by manipulating the relative frequencies with which the stimuli were presented both in the preshift and postshift stimulus series.

EXPERIMENT I

With the aim of establishing two levels of discriminability of stimulus shift, two different preshift distribu-

[1] This research was carried out while the author held a fellowship from the National Science Foundation. Experiments II, III, and IV were reported in a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of California. The author is indebted to Dr. Leo Postman, thesis director.

[2] Now at Swarthmore College.

tions of stimuli were selected: one was positively skewed while the other was negatively skewed. The postshift series, the same under both conditions, was a restricted range of stimuli drawn from the lower half of the preshift range. It was assumed that the stimulus *shift* would be more readily discriminated when the postshift stimuli were those which appeared least frequently in the preshift distribution, and that this would result in a more rapid readjustment of the judgment scale to the postshift stimuli.

A second variable explored in Exp. I was the amount of experience *Ss* had with the preshift distributions. Both Johnson (**4**) and Tresselt (**8**) had found that the rate of readjustment to the postshift range was an inverse function of the number of presentations of the preshift stimuli. This variable was introduced into the present experiment because of the surprisingly small amount of shift found with the initial groups which had been exposed to an arbitrarily selected number of preshift presentations.

Method

Subjects.—The *Ss* were 112 students in elementary psychology at the University of California. They were randomly divided into seven groups with from 12 to 20 *Ss* in each.

Materials.—The stimuli were nine square, black cards with the following widths: 2, 2½, 3, 3½, 4½, 5½, 6½, 7½, and 9 in.

Procedure.—Each group (the respective groups were run at different hours) was given response sheets and was read the following instructions:

"Your task in this experiment is to judge the size of different square cards. You will be presented with a series of cards like this [the first card was shown]. As each card is presented, you make a judgment of how large or small it appears to you. Record your response in terms of one of the five size categories listed at the top of the page. Use the abbreviations shown: if the card looks very large to you, write VL; large—L; medium—M; small—S; and very small—VS. Do not go back and change your responses after new stimuli have been presented. This experiment concerns subjective impressions only; and since you are to judge how large the card looks to *you* on *that* presentation, there can, of course, be no right or wrong answers. Please respond to every card even if the decision seems difficult with the first ones."

After these instructions, the stimuli were displayed manually to *Ss*, being held one at a time before a white, 28 × 44-in. backdrop. The *E* called out the number of each of the successive presentations to help *Ss* record their judgments in the proper place. For each of the six experimental groups there were a total of approximately 172 stimulus presentations (plus or minus one to even out the series). The three levels of preshift experience consisted of 44, 88, and 132 presentations, respectively. At each of these levels, one group (hereafter designated N) was presented with stimuli from the negatively skewed distribution in which each of the five largest stimuli appeared eight times as often as each of the four smallest stimuli. The order was quasi-random, i.e., the smallest cards appeared on presentations 9, 15, 30, 37, etc.; *each* of the five largest cards appeared equally often during the remaining presentations in a block pattern such as was used earlier by Parducci (**6**). For the three groups judging stimuli from the positively skewed distribution, the P groups, each of the five smallest stimuli appeared eight times as often as each of the four largest stimuli. For all six experimental groups, the postshift series consisted of the five smallest

TABLE 1

PRESENTATION OF THE STIMULI IN EXP. I–IV

Group	Preshift Series*	Postshift Series*
P–44 (PR)	1–5[8]; 6–9[1]	1–5[26]
P–88	1–5[16]; 6–9[2]	1–5[17]
P–132	1–5[24]; 6–9[3]	1–5[8]
N–44 (NR)	1–4[1]; 5–9[8]	1–5[26]
N–88	1–4[2]; 5–9[16]	1–5[17]
N–132	1–4[3]; 5–9[24]	1–5[8]
N–88A	1–4[2]; 5–9[16]	1–5[17]
Control	1–5[20]	
PE	1–5[9]	1–5[24]; 6–9[3]
PE–pitch	1–5[9]	1–5[24]; 6–9[3]
PR–pitch	1–5[8]; 6–9[1]	1–5[26]
NE	1–5[9]	1–4[3]; 5–9[24]
NE–color	1–5[9]—Blue	1–4[3]; 5–9[24]
NR–color	1–4[1]; 5–9[8]	1–5[26]—Blue

* Unbracketed numerals represent stimulus values (assigned in order of increasing size or pitch); adjacent bracketed numerals represent respective frequencies of presentation.

stimuli presented repeatedly in quasi-random order as in the preshift P series but without the four largest stimuli. A control group judged only this postshift series, receiving 100 presentations in all. The respective stimulus series are summarized in Table 1.

The specific hypotheses were: (*a*) that the N groups would readjust to the postshift series more quickly than the P groups, and (*b*) that the rate of shift from both preshift distributions would decrease as a function of amount of experience with these distributions.

TABLE 2

EXP. I: MEAN SHIFT IN JUDGMENT

Group	*N*	Mean	*SD*
P–44	17	.08	.19
P–88	15	.08	.25
P–132	17	.02	.18
N–44	16	.42	.36
N–88	12	.72	.58
N–132	15	.35	.44
N–88A	20	.78	.37

Results

The judgments were converted to numerical form: very large = 5, large = 4, medium = 3, small = 2, very small = 1. To simplify the calculations, only the judgments of the middle stimulus were used. This was the one stimulus which appeared with maximum frequency in both the P and N preshift series, and the judgments of it were the ones most affected by the stimulus shift.

The intention had been to compare the different experimental groups on the basis of the number of postshift presentations necessary before their judgments of the middle stimulus were the same as the mean judgment of it made by the control group. This proved to be impossible. The control group on every trial judged it to be between "large" and "very large," the tabulated mean for its last eight judgments of this stimulus being 4.6. Of the six experimental groups, none ever approached this average; Group P–44, which came the closest, at no time exceeded 3.6, a full judgment category lower than the control's average judgment.

It was this failure to readjust the judgment scale to the postshift series that had suggested exposing groups to different amounts of preshift experience (the two 88-presentation groups having been run first). In order to evaluate the effects of different amounts of preshift experience and also to compare the effects of the two preshift distributions, a measure of each *S*'s judgment *shift* was calculated by subtracting his mean judgment of the last three preshift presentations of the middle stimulus from the mean of his first eight postshift judgments of this same stimulus. The group means and *SD*'s are presented in Table 2. An analysis of variance performed upon these scores (with all groups reduced to $N = 12$) indicated that the effect of the preshift distributions (i.e., assumed discriminability of the shifts) was statistically significant ($P < .001$); the effect of the different amounts of preshift experience was less so ($P < .05$); and the interaction was not significant. A corresponding analysis using the relatively distribution-free Mood Median Test (**5**) yielded similar probabilities except that the effect of the amount of preshift experience appeared less significant ($P < .20$). Use of the Mood test was suggested by the marked heterogeneity of variance (Bartlett's test: $P < .01$) which reflected the fact that over half of the P group *Ss* made no shift at all in their judgments. An analysis of variance performed upon the three N group shifts was significant ($P < .01$). The apparently curvilinear relationship between the amount of shift and the amount of preshift experience was, however, a completely unexpected characteristic of these results. Group N–88, not Group N–44, showed the greatest shift. Additional evidence that this inversion was not an artifact of sampling is provided by the per-

TABLE 3

EXP. I: MEAN PRESHIFT JUDGMENTS

Group*	*N*	Mean	*SD*
P	49	3.16	.42
N	43	2.12	.38

* The three P and the three N groups combined.

formance of an additional group of 20 *Ss* (Group N–88A) run under the same conditions as Group N–88 with the exception that it was presented with *each* of the nine stimuli used in the preshift distribution during the first nine presentations (instead of having the four smallest stimuli scattered randomly over the first 44 presentations, the remaining four presentations of these smallest stimuli were randomly scattered over the entire 88-presentation series). This group was added under the assumption that an initial experience with the total range might *anchor* the judgment scale, minimizing subsequent changes in judgment. The amount of shift by the new group, however, was even larger than that by Group N–88 (see Table 2). Both of these groups showed significantly more shift ($P < .01$) than either Group N–44 or Group N–132.

An incidental result of this experiment (see Table 3), not relevant to the hypotheses under consideration but of some significance for the study of the formation of judgment scales, was that the mean of the last three preshift judgments of the middle stimulus by *Ss* in the three P groups was significantly larger than the corresponding means for the three N groups ($P < .0001$).

Discussion

This last, incidental finding supports the weighted-mean approach taken by Helson (2) and Johnson (3). The experimental groups all judged stimuli covering the same range; the only difference between their treatments was in the distribution, or relative frequency, of the stimuli within that range. In accordance with the weighted-mean approach, the N series (with the larger mean stimulus value) evoked "smaller" judgments of the middle stimulus than did the P series. These results suggest limitations for theories that stress the range of the stimuli (9).

The greater judgment shifts by the N than by the P groups is also consistent with the weighted-mean approach. This is because the mean of the postshift stimuli is closer in value to the mean of the positively skewed preshift series. The differences in judgment associated with *amount* of preshift experience are not consistent, however, with either this approach or with the experimental findings of Johnson (4) and Tresselt (8). It should be noted that there are prominant differences between their experiments and the present one. Johnson's *Ss* judged the pitch of tones while Tresselt's judged lifted weights. The stimulus ranges of Johnson's pre- and postshift distributions did not overlap. Tresselt's overlapped, but her shift involved an *extension* rather than a *restriction* of the stimulus range. Whatever the crucial differences, however, the present data suggest that the rigidity of the judgment scale is not a simple, increasing function of the amount of preshift experience.

The most surprising outcome of this experiment was the smallness of the judgment shifts. The three P groups shifted hardly at all, and none of the groups (not even Groups N–44 and P–44 that had 26 repetitions of the postshift series) ever approached the level of the control group. These results are quite different from Johnson's (4) in which the control level had been reached after only three repetitions of the postshift series. While it had been anticipated that the P groups would be slow to shift (because they would not immediately recognize that the large stimuli were no longer appearing), such considerations would

hardly apply to the N groups. The results of this experiment thus raise doubts concerning the conceptualization of the judgment process as one in which *S* fits his scale to the distribution of stimuli he is judging. Here we appear to have found a fixation of scale which leaves *S* relatively insensitive to restriction of the range of the stimuli. It would seem important to study this situation further, attempting to single out those of its characteristics which might account for this unexpected failure to find greater shift in judgment.

Experiment II

To determine whether the failure to shift in Exp. I was due to the range being restricted rather than extended, the *direction* of the stimulus shift was reversed. A new group of *S*s was presented with a preshift series corresponding to the former postshift series, and with a postshift series corresponding to the former preshift series. Thus for this new group, the stimulus shift involved an *extension* rather than a *restriction* of the total range of stimulus values. To make the results more comparable with Johnson's and also to increase their generality, two additional groups were asked to judge the pitch of a series of tones. The respective frequencies of these tones corresponded to the different sizes presented to the card-judging groups.

Method

Subjects.—Three groups, totaling 46 *S*s, were drawn from the introductory psychology courses. The judgments made by Group P–44 in Exp. I were also used in this experiment.

Materials.—The two card-judging groups were presented with the stimuli used in Exp. I. The other two groups were presented with pure tones. These had been recorded on a sound tape, an oscillator and a voice microphone (through which the number of each succeeding judgment was announced before the tone sounded) both feeding into the tape recorder. Separate tone series were recorded for the two groups. The frequencies of the respective tones were 200, 400, 600, 800, 1000, 1200, 1400, 1700, and 2000 cps.

Procedure.—The new card-judging group (hereafter called Group PE) received the same instructions as had been given to Group P–44 (hereafter called Group PR) for which the range had been restricted. Appropriate changes were made in the instructions to the two pitch-judging groups (PE–pitch and PR–pitch). Their judgments were to be made in terms of a five-category pitch scale: very high, high, medium, low, and very low. The preshift series for both the PE groups consisted of 45 randomized presentations, nine each of the five smallest (or lowest pitched) stimuli. The postshift series for these two groups was composed of 132 presentations of the preshift P series, following the same order of presentation (i.e., the preshift series presented to Group P–132 of Exp. I). The third new group, PR–pitch, judged tones whose successive stimulus values followed the same order as the cards presented to its counterpart, the former P–44 group (see Table 1).

Results

The judgments were again converted to numerical form, just as in Exp. I. In this and the following experiments, the difference between each *S*'s mean judgment of the last three preshift presentations of the middle stimulus and the mean of his first eight postshift judgments of this same stimulus was computed. This provided a measure of his *immediate* shift. The measure of each *S*'s *total* shift was the difference between his preshift mean and the mean of his *last* eight judgments of the middle stimulus. The sign of this difference for each group was kept positive by subtracting, under each condition, the mean judgment for the wide range from the corresponding mean for the narrow range (this yielded small negative scores for certain *S*s but never for the group means).

The results for Exp. II are presented in Table 4. It should first be noted that the magnitude of the

TABLE 4

EXP. II: MEAN SHIFT IN JUDGMENT

Group	*N*	Immediate Shift		Total Shift	
		Mean	*SD*	Mean	*SD*
Cards:					
PE	12	.72	.57	1.14	.44
PR	17	.08	.19	.19	.35
Pitch:					
PE–p	14	.52	.28	.70	.37
PR–p	20	.04	.40	.43	.44

TABLE 5

EXP. II, III, IV: ANALYSES OF VARIANCE OF JUDGMENT SHIFTS

Source	Immediate Shift†		Total Shift	
	Mean Square	*F*	Mean Square	*F*
Experiment II				
Direction of shift	.164	27.33**	4.43	24.61**
Nature of stimuli	.015	2.50	.20	1.11
Interaction	.001		1.17	6.50*
Within	.006		.18	
Experiment III				
Direction of shift	.251	43.65**	13.88	45.06**
Discriminability	.108	18.78**	6.36	20.65**
Interaction	.017	2.94	.29	
Within	.006		.31	
Experiment IV				
Direction of shift	12.70	33.42**	10.39	24.74**
Color change	1.28	3.37	2.51	5.98*
Interaction	.86	2.27	1.09	2.60
Within	.38		.42	

Note.—All groups reduced to $N = 12$ by random selection. For all main effects and interactions, $df = 1$; $df = 44$ for each within-cell variance estimate.
* Significant at the .05 level.
** Significant at the .001 level.
† Based upon logarithmic transformations of the difference scores in Exp. II and III, but not in Exp. IV.

judgment shifts was much greater when the range of the stimuli was extended rather than restricted; in fact, both PE groups reached the preshift level of their corresponding PR groups, i.e., ended up by judging the middle stimulus when it occurred in the total range with the same average category as would have been applied to it if the total range had been presented from the beginning. An analysis of variance (see Table 5 for this and all subsequent analyses of variance) performed upon the total-shift scores revealed a significant interaction ($P < .05$) between the direction of stimulus shift (extension or restriction of range) and the nature of the stimulus materials (cards or tones). There was no indication of interaction, however, in a corresponding analysis of the immediate-shift scores (transformed logarithmically to reduce heterogeneity of variance due to almost complete absence of shift by the groups for whom the range was restricted). With both analyses, the direction-of-shift effect was highly significant ($P < .001$); but it was not significant at the .05 level when tested against the interaction in the total-shift analysis. Three of the *t* tests between the means of the corresponding extension and restriction groups yielded differences significant at the .001 level; the fourth, the total-shift comparison for the two pitch groups, was significant at only the .10 level.

Discussion

The results indicate that with these stimuli an extension of the range evokes a greater shift in judgment than does a restriction of the range. While this effect was not so pronounced with the total-shift measure of the judgments of pitch, there are grounds for regarding the immediate-shift scores as more informative in this respect. This is because the mean preshift judgments by the two pitch groups were only slightly more than one-half a category apart, most *S*s judging the middle tone to be either *high* or *very high*. This may have been due

to a contrast effect produced by the interpolated voice of *E* (calling out the presentation numbers). Whatever its basis, it placed a limit upon the amount of shift the pitch groups could be expected to make; and this limit was reached by Group PE–pitch before it was two-thirds of the way through its postshift series. Group PR–pitch never reached its expected limit, the preshift level of Group PE–pitch.

These results, while they perhaps have a certain common-sense obviousness, are not readily subsumed under the more general judgment principles presented by Helson (**2**), Johnson (**4**), and Volkmann (**9**). Nor would a differential weighting for expected *recency* effects (**4, 6**) handle the differences in shift. Some rationale would be required for giving greater weighting to the wider distribution than to the narrower one. Until more is known about the effect of the direction of stimulus shift, there could be little basis for such differential weighting.

Experiment III

A possible explanation for the smaller judgment shifts following restriction of the stimulus range is that *Ss* could not *discriminate* the stimulus shift. If the restriction of the range were easier to discriminate, as it was assumed to be for the N groups in Exp. I, the effect of the direction of shift should be less pronounced than it was in Exp. II. To test this hypothesis, an additional group of 12 *Ss*, Group NE, was run under the same conditions as Group PE except that it judged as its postshift series the preshift series presented to Group N–132 in Exp. I (see Table 1). The difference in shift between Groups NE and NR could thus be compared with the difference between Groups PE and PR; if the smaller shifts following restriction of range are due to the difficulty of discriminating this restriction, the two N groups should be more alike in the magnitude of their shifts than the two P groups. This discriminability hypothesis would be supported by a significant interaction between the effects of the direction of the stimulus shift and the discriminability of the shift.

TABLE 6

Exp. III and IV: Mean Judgment Shift

Group	*N*	Immediate Shift		Total Shift	
		Mean	*SD*	Mean	*SD*
Exp. III:					
NE	12	1.70	.82	2.02	.79
NR	16	.42	.36	.77	.51
Exp. IV:					
NE—color	12	1.76	.64	2.18	.58
NR—color	15	.90	.54	1.37	.73

Results and Discussion

The scores were tabulated just as in Exp. II. The means and *SD*'s of Groups NE and NR are presented in the top portion of Table 6 (for Exp. III, they are to be compared with the means of Groups PE and PR presented in the top half of Table 4). Analyses of variance (Table 5) were performed upon both the raw total-shift scores and upon logarithmic transformations of the immediate-shift scores (transformed to reduce heterogeneity of variance). In neither analysis did the interaction approach significance (in both, $P > .05$); but the two main effects, discriminability of shift (P or N) and direction of shift (E or R), were highly significant ($P < .001$ for all four F ratios). Thus while the results give added support to the previously observed importance of the direction of the stimulus shift, they provide no basis for ascribing this effect to *Ss*' inability to discriminate the restriction of the stimulus range.

Of the 16 *Ss* in Group NR, 11 showed an "immediate shift" with the dropping out of the four largest stimuli and 13 had shifted by the end of the postshift series (total-shift measure). No *S* in this group decreased his judgment of the middle stimulus after the shift. This indicates that Group NR was responding systematically to the shift and is consistent with the assumption that the shift was more readily discriminable for Group NR than for Group PR; for of the 17 *Ss* in Group PR, 5 *Ss* showed no "total shift" at all, and 4 *Ss* actually shifted in the opposite direction.

The tendency for the N groups to shift more than the P groups was already discussed in conjunction with Exp. I where it was noted that the discriminability effect is consistent with judgment theories emphasizing the respective frequencies of the different stimulus values (**2, 4**).

Experiment IV

In a previous experiment (**6**), an interpolated series of larger cards was found to have less of an anchoring effect upon rejudgments of the original series of cards when the two series were different in color than when they were the same color. It seemed unlikely that the effect of this unjudged dimension could have been based upon any difference in the discriminability of the shifts between the two series since all *Ss* responded to these shifts (with corresponding change in judgment) even when both series were the same color. Perhaps the change in color indicated to *Ss* that there were two *separate* series which were to be judged by different standards. Following this line of speculation, it might be argued that the direction-of-shift effect in the present experiments was due to a difference in the information provided by the two directions of shift. When the range of the stimuli was extended, *Ss* recognized at once that the postshift stimuli formed a new series. When the range was restricted, *Ss* might have discriminated the stimulus shift, recognized that some of the stimuli were not being shown, and yet not have attributed the shift to a permanent change in the stimulus series. They might thus have saved their extreme (e.g., largest) judgment categories for the missing members of the preshift series.

To test this possibility, two new groups (NE–color with 12 *Ss* and NR–color with 15 *Ss*) were exposed to the same stimuli as Groups NE and NR in Exp. III (see Table 1); however, the preshift stimuli for Group NE–color and the postshift stimuli for Group NR–color were blue rather than black in color. The hypothesis was that the introduction of this color shift should produce greater shifts in judgment. The *increase* in shift should be greater following restriction of range since, even without the change in color, *Ss* had presumably recognized the permanent change in the distribution when the range was extended.

Results

The means and *SD*'s of the two new groups are presented in the bottom of Table 6. In accordance with the hypothesis, Group NR–color shifted more than Group NR on both the immediate- and total-shift measures ($P < .01$ and .02, respectively). The slightly greater shifts by Group NE–color than by Group NE did not approach significance on either measure; this also is consistent with the experimental hypothesis. However, analyses of variance performed on both the immediate- and total-shift scores (see Table 5) indicated that neither interaction was significant. This means that the greater effect of color change when the stimulus range is restricted rather than extended is

not statistically significant when all four groups are compared in one analysis.

The direction-of-shift effect was again significant ($P < .001$ for both immediate and total shifts), and the effect of color change was significant at the .05 level for the total-shift but not for the immediate-shift analysis.

Discussion

The results of Exp. IV are consistent with the assumption that introduction of a color change would serve to further isolate the pre- and postshift stimulus series. Some of the conditions determining the degree of this isolation in all four experiments might be tied together by means of an analogy to random sampling from a hypothetical population. Assume that the hypothetical population of stimuli for each group of *S*s is the actual distribution of stimuli in the preshift series. Each postshift series could then be evaluated in terms of the probability that it is a chance sampling from this hypothetical population. For the groups whose stimulus shifts involve an *extension* of the range of the preshift series, the postshift series are unlikely samples from the hypothetical population because they include stimuli not present in that population. For the groups whose stimulus shifts involve a *restriction* of the range, the postshift stimuli could all be drawn from the hypothetical population. The respective probabilities that they are such samples, however, are different for the different experimental conditions. For the PR groups, the postshift "samples" are almost identical with the preshift population—the only difference being the omission of an occasional large stimulus. For the NR groups, however, these large stimuli occur most of the time during the preshift series; and thus the probability of the narrow, postshift series being random samples from the "population" is considerably less. This difference in the respective probabilities for the PR and NR groups is paralleled in Exp. I by the highly significant difference in amount of judgment shift as a function of the shape of the preshift distribution.

Although none of the groups in these experiments completely readjusted their judgment scales to the restricted postshift range, such readjustment might have occurred under less ambiguous circumstances. Estes and Straughan (1) found that in a Humphreys-type, "verbal conditioning" situation *S*s quickly learned to respond in accordance with the actual postshift probability of reinforcement, .30, when shifted from a preshift series with a markedly different probability, .85. However, other *S*s who were shifted from a series in which the reinforcement probability was only .50 did not learn to respond in accordance with the .30 postshift probability. This difference fits in well with our random sampling analogy since the .30 series is a more likely sample from a .50 than from a .85 population.

The analogy is less clear for Group NR–color of our Exp. IV. Although its postshift series was from a new population in respect to color, it could still have been a sample from the old population in respect to size. Here, the effect of the color change might have been to call attention to the possibility of a corresponding size change. The NR–color *S*s would therefore have been less ready to attribute the observed changes in the distribution of stimulus sizes to the temporary vagaries of random sampling, instead shifting their size judgments in accordance with the new distribution.

The rationale behind this sampling analogy is that *S*s in a judgment situation of this kind tend to fit their response scales to the total range of the stimuli which they have learned to expect. If either earlier presentations or the experimental instructions (6) lead them to expect stimuli outside the presented range, they will not use all of the response categories.

The decision as to which category is to be applied to each specific stimulus depends, at least in part, upon the relative frequency with which the different stimuli have been presented. Thus

in the preshift series of Exp. I, it was found that the judgments of the middle stimulus depended upon whether it appeared in a positively or negatively skewed series. It seems likely that there are rules or language habits governing the relative frequency with which the different judgment categories are to be used, and *Ss* can apply these rules only insofar as they have learned the relative probability of appearance of the respective stimuli they are judging. The different effects upon judgment of the stimulus shifts utilized in the preceding experiments might thus depend upon the kind of information they impart to *Ss* concerning the probable distribution of the new series.

This analogy can hardly be said to *explain* the results of these experiments. It does, however, suggest the desirability of further research in which the relative frequencies of the different pre- and postshift stimuli would be systematically varied, and it suggests that the basic problem in describing judgmental behavior may be the determination of the principles of probability learning.

Summary

Four experiments were presented investigating the effects upon judgment of different kinds of shifts in the range and distribution of the stimulus series.

In Exp. I, *Ss* were presented with a series of square cards and instructed to judge them in terms of five alternative size categories. In the course of the experiment, they were shifted to a new series from which the larger members of the original series had been dropped. Contrary to expectations, many *Ss* did not readjust their judgment scales to the narrower range. This resistance to shift was greater for the groups whose preshift stimuli were closer in mean value to the mean of the postshift stimuli, and it also appeared to be a curvilinear function of the number of presentations of the original series. Judgment shift was greatest following an intermediate amount of experience with this series, the resistance to change being greater with either more or less preshift exposure.

In Exp. II, a greater shift in judgment was found when *Ss* were presented first with a narrow range of stimuli and then shifted to a wider range than when the stimulus shift was in the opposite direction. This direction-of-shift effect was found both for judgments of the size of square cards and the pitch of a series of tones, and its magnitude was not altered by an increase in the discriminability of the stimulus shift (Exp. III). The amount of shift following a narrowing of the stimulus range was increased, however, by a simultaneous change in the color of the stimuli (Exp. IV).

An additional, incidental finding in Exp. I was that the same stimulus was judged larger when the distribution of stimuli in which it appeared was positively skewed than when it was negatively skewed, even though both distributions had the same range.

References

1. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
2. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
3. Johnson, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, **34**, 425–436.
4. Johnson, D. M. Learning function for a change in the scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851–859.
5. Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
6. Parducci, A. Learning variables in the judgment of single stimuli. *J. exp. Psychol.*, 1954, **48**, 24–30.
7. Postman, L., & Miller, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58**, 43–53.
8. Tresselt, M. E. The influence of amount of practice upon the formation of a scale of judgment. *J. exp. Psychol.*, 1947, **37**, 251–260.
9. Volkmann, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer and M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951.

(Received April 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

TIME UNCERTAINTY IN SIMPLE REACTION TIME[1]

EDMUND T. KLEMMER

Operational Applications Laboratory, Air Force Cambridge Research Center

A recent review by Teichner (**3**) describes the state of our knowledge concerning simple reaction time. Although much work has been done in this field, it seems that the most basic variable in simple RT tests has been largely overlooked—the time variable. By definition, the only uncertainty *S* has in a simple RT test is time uncertainty. He knows exactly what the stimulus will be and what the response will be. He does not know when the stimulus will be presented.

The effects of changes in time of stimulus occurrence are usually discussed under the factors of readiness, set, or attention. These classifications lead to a concentration upon the length of the foreperiod with little attention paid to foreperiod variability (**6**, p. 314).

The *S*'s time uncertainty is the result of both his own imperfect time-keeping ability and the clock-time variability of the stimuli. The first factor varies with the length of constant foreperiods or interstimulus intervals and the second is defined by the foreperiod variability. The tests of the present study are so designed that the effect of these two factors can be considered separately. Very simply, what we did was run a simple reaction-time experiment with careful control of foreperiod length and variability and control of stimulus repetition rate.

[1] This research was performed at the Operational Applications Laboratory, Bolling Air Force Base, Washington, D. C. in support of Project 7682. Reprints of this paper will be distributed as Air Force Cambridge Research Center Report TR-55-2.

METHOD

Apparatus.—The apparatus consisted of an NE 51 neon stimulus bulb, a telegraph-response key, and a warning-click device. The .02-sec. stimulus was clearly visible in the dimly lighted room. In another room a teletype tape programmer presented 11 different lengths of foreperiod in random order. The total range of foreperiods could be varied from 0 to 8 sec. The minimum foreperiod could be varied independently of the range. The over-all cycling rate was adjustable down to one cycle every 12 sec.

Each reaction time was automatically encoded into cycle counts of a 50-cps oscillator and punched on teletype tape. In addition, RT's were accumulated on a Standard Electric timer.

Procedure.—This study consists of two test series. In the first series the warning click occurred at 11-sec. intervals and the foreperiod variability and mean foreperiod changed between tests. These tests will be called the Foreperiod Series. The second series had no warning click and no variability in time of stimulus occurrence during each run but the interstimulus interval was changed between tests. This will be called the Interval Series.

The values of mean foreperiod and foreperiod variability for the Foreperiod Series and the intervals between stimuli for the Interval Series are shown with the data in Tables 1 and 2. The apparatus was programmed to produce rectangular distributions of foreperiods. Foreperiod variability is reported in terms of the range of the distributions.

The *S*s were six university students who were paid for their services. Each *S* was given extensive training on the apparatus over the range of conditions used in the test series. Every *S* took all of the tests shown in Tables 1 and 2. The Foreperiod Series was given before the Interval Series. The test runs within each series were randomized. In addition, two of the *S*s were given 10 additional consecutive runs on Test M because of particular results that were noted in this test.

An attempt was made to run a test in the Interval Series with stimuli repeated every .5 sec. This test was dropped when it was discovered that no *S* could consistently keep in phase with stimuli presented at that rate.

In all tests *S* was instructed to respond as soon as possible after the stimulus light but never before. If a response occurred before the stimulus in any run, the run was halted and started over. This was sufficient negative reinforcement to keep the number of early responses negligibly small.

Results[2]

Foreperiod Series

The mean RT's for each of the seven Foreperiod Series tests are shown in Table 1, and the means for each foreperiod in these seven tests are plotted in Fig. 1. Let us look first at these mean RT's. Figure 1 shows that mean RT increases with mean foreperiod in the range of these data, but that the amount of increase is much less for the tests with 2-sec. variability (S, Q, P) than for tests with no variability (M, L, K). In addition, mean RT increases with the variability of the foreperiod where foreperiod variability is represented by the spread of points of the same test along the abscissa. The effect of foreperiod variability upon mean RT

[2] The complete set of 72 mean RT's, *SD*'s, and ranges for the 6 *S*s and 12 tests has been deposited with the American Documentation Institute. Order Document No. 4765, remitting $1.25 for microfilm or $1.25 for photocopies.

TABLE 1

Description of Foreperiod Series Tests

(Constant 11-sec. spacing of warning signals; 330 RT's for each of six *S*s)

Test	Mean Foreperiod (Sec.)	Foreperiod Range (Sec.)	Mean RT (Sec.)
K	8.25	0	.269
L	4.25	0	.252
M	.25	0	.209
P	7.25	2	.272
Q	4.25	2	.259
S	1.25	2	.259
T	4.25	8	.281

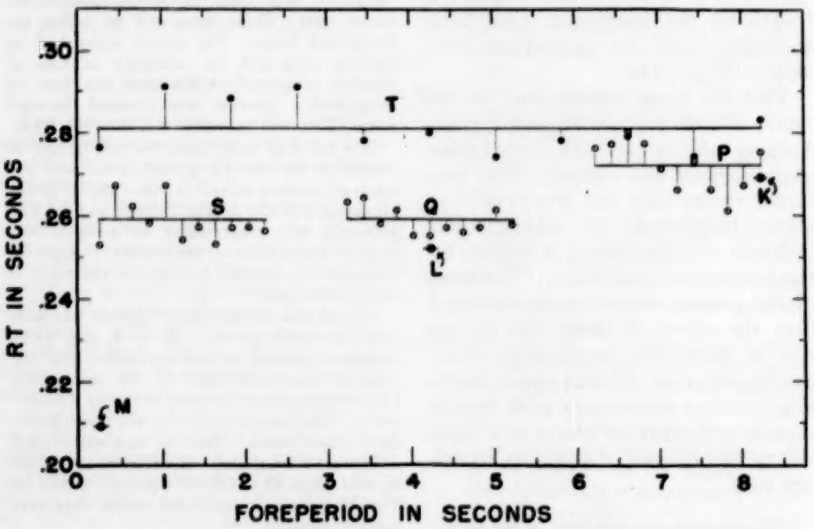

Fig. 1. The influence of foreperiod and foreperiod variability upon RT. Combined data of six *S*s. Each point is the mean of 180 RT's in tests T, S, Q, P, and 1980 RT's in tests M, L, K. Tests T, S, Q, P have randomly varying foreperiods within each run. Tests M, L, K have constant foreperiods for all stimulus presentations. Straight lines are drawn through the mean RT values for each test.

seems greater for tests with smaller mean foreperiod, but this conclusion must be tentative since only tests L, Q, and T have the same mean foreperiod.

Additional comparison among tests are made possible by selecting three specific foreperiods, each of which is common to three tests. Reaction times to each of these foreperiods in all tests in which they occur are plotted against foreperiod variability in Fig. 2. This figure shows that RT to each selected foreperiod does increase with the range of foreperiods within which it is embedded, but that this increase is much less if the foreperiod is large than if the foreperiod is small. Thus, at a .25-sec. foreperiod, increasing the range of foreperiods from a 0 to 8 sec. adds 67 msec. to RT while at a 8.25-sec. foreperiod the same increase in foreperiod variability only adds 14 msec. to RT.

We shall consider next the RT differences within tests with variable foreperiods. The influence of immediate foreperiod upon RT within each test may be seen in the scatter of points about the mean RT lines in Fig. 1. Short foreperiods generally

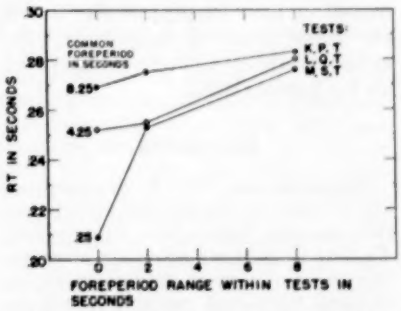

FIG. 2. The effect of foreperiod variability upon RT for each of three selected foreperiods common to several tests. Combined data of six *Ss*. Zero variability points are means of 1980 RT's each. Other points are means of 180 RT's each.

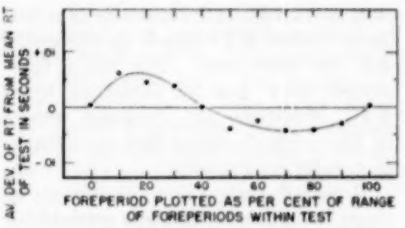

FIG. 3. The influence of foreperiod within tests with variable foreperiod upon RT. Combined data of six *Ss* and four tests: S, Q, P, and T. Each point is the mean of 720 RT's.

result in higher RT's than long foreperiods within each test. This relation is statistically significant using a simple chi-square test of contingency.

There is actually a suggestion of a sine-wave relation between foreperiod and RT rather than simple regression. This may be best seen in Fig. 3 which plots the deviation from average RT against foreperiod range (standardized to one unit) for all four variable foreperiod tests combined. Note that the amplitude of this effect (± .005 sec.) is small.

Pattern of foreperiods.—In order to determine the effect upon RT of the pattern among foreperiods within each run, the RT's of the variable foreperiod tests were tabulated according to the triad of foreperiods which preceded each RT. For this analysis the foreperiods were encoded in binary fashion relative to the mean foreperiod. The triad foreperiod analysis gave RT's differing only a few milliseconds from the mean RT in each test, but some consistencies are noted. In all tests the highest RT's result from a short foreperiod preceded by a long foreperiod (6 msec. average increase). The reverse, a long foreperiod preceded by a short foreperiod, results in the shortest RT's in all tests (5 msec. average decrease). Two long foreperiods in a

row or two short foreperiods in a row gave rise to RT's equal to the mean RT for that test. The third foreperiod back has no influence upon RT. Woodrow (5) obtained results in the same direction but apparently of greater magnitude.

Repeated tests—The situation at short foreperiods and no variability is further complicated by the fact that these tests give lower RT's when run alone than when run in company with other tests. It was noted during the runs of Test M of the Foreperiod Series (.25-sec. foreperiod, no variability) that *S*s were producing more very low RT's than in other tests of the series. In order to investigate this effect further, two *S*s were given 10 consecutive runs of Test M alone. The last 6 of these were combined and compared to the original 6 runs for the same *S*s. The mean RT dropped from .176 sec. in the original tests to .109 sec. for the repeated tests. When Test T (8-sec. range of foreperiod) was given to four *S*s 10 or more consecutive times, no consistent shift in RT distribution resulted. It is evident that the large reduction in RT that arises from repeating Test M is due to its low time uncertainty and not merely to a general practice effect.

Interval Series

The mean RT's for each of the five fixed Interval Series tests are shown in Table 2, and are plotted against the interval in Fig. 4. The time interval between stimuli has been plotted on a log scale and a straight line fitted to the points by eye. The increase in RT in going from a 1-sec. interval to an 8-sec. interval with no variability is 65 msec. This increase is just about the same as the increase which results from adding 8 sec. of foreperiod variability to the .25-sec. foreperiod (see Fig. 3). This would suggest that changes in RT due to clock-time uncertainty and due to *S*'s imperfect time keeping ability are of the same order of magnitude. No stronger statement seems warranted on the present data.

TABLE 2

DESCRIPTION OF INTERVAL SERIES TESTS

(No warning signal; no variability; 300 RT's for each of six *S*s)

Test	Interstimulus Interval (Sec.)	Mean RT (Sec.)
B	1	.203
C	2	.223
D	4	.233
E	8	.256
F	12	.268

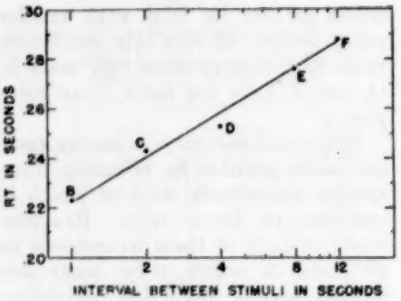

FIG. 4. The influence of time between stimuli in tests with no warning click and no variability. Combined data of six *S*s. Each point is the mean of 1800 RT's. Letters alongside points designate tests described in Table 2.

DISCUSSION

Foreperiod and RT.—It seems reasonable in view of the above results to say that reaction time varies directly with *S*'s time uncertainty about the stimulus. Increasing the mean foreperiod or increasing foreperiod variability will increase *S*'s uncertainty about the time of occurrence of the stimulus and thus increase RT.

The positive relation between mean foreperiod and RT will hold only down to some optimum value of foreperiod,

below which *S* cannot make full use of the warning signal in time. This optimum mean foreperiod is small for low variability tests. The .25-sec. constant foreperiod of Test M gave by far the fastest RT of any tests (Fig. 1). Although the present study does not provide data for low variability foreperiods with means around 1 or 2 sec., several pilot study tests do show that there is no minimum above 1 sec. in the RT vs. foreperiod curve. Specifically four pilot tests show that with .5 sec. of foreperiod variability or less, lower RT's are obtained with mean foreperiods less than 1 sec. than with mean foreperiods between 1 and 4 sec. The tests of the present study with a 2-sec. range of foreperiods indicated no effect due to mean foreperiod below 4.25 sec. Thus it appears that there is a critical optimum mean foreperiod only when the foreperiod variability is small, and that for small variability the optimum foreperiod is less than 1 sec.

These findings are not in harmony with previous statements about optimum foreperiod which are summarized by Teichner in his conclusion that "the optimum foreperiod of RT may be thought of as lying in the range between approximately 1.5 and 8.0 sec." (**3**, p. 144). His conclusion is based largely on data from Breitwieser, Woodrow, and Telford, although different procedures were involved. Breitwieser (**1**) found a 1-msec. effect favoring foreperiods of 2 or 3 sec. when foreperiods were randomly mixed. Woodrow (**5**), using a similar procedure, found small differences indicating that 3 sec. was the least favorable foreperiod, and the present study indicates that the optimum foreperiod in a test with randomly ordered foreperiods is dependent upon the range of variation of the foreperiods (Fig. 1 and 3). All of these results show that the important effect of randomly varying foreperiods is not the small differences of RT within the test but the fact that every one of the foreperiods run in random order leads to a higher RT than any one of them given repeatedly by itself.

Optimum foreperiod has meaning in tests with fixed foreperiod. In Woodrow's tests with constant foreperiods of from 1 to 36 sec., he found that 2 sec. gave the lowest RT. The present results, together with pilot study data, indicate an optimum fixed foreperiod less than 1 sec. It seemed that a likely contributor to this disagreement might be the difference in time uncertainty of the warning signal in the two procedures. The present results are based on perfectly regular warning signals whereas Woodrow used a procedure which introduced uncontrolled variability in the time of occurrence of the warning signal. In order to test this hypothesis, we ran four *S*s in tests with time-variable warning signals and constant foreperiods, and compared mean RT's with alternately given tests with the same constant foreperiod and regularly occurring warning signals. Eight seconds of time variability in the warning signals produced RT's differing irregularly by only a few milliseconds from the tests with no variability. Thus the hypothesis that the difference between Woodrow's results and the present study are due to time variability in the warning signal must be rejected.

Telford (**4**) used no warning signals but his procedure can be interpreted in foreperiod terms by treating each stimulus as a warning signal for the next stimulus. His tests then resemble a randomly variable foreperiod test such as Test T of the present study. Unlike the results of Breitwieser, Woodrow, or this study, however, Telford reported a large effect due to interval (foreperiod) with random series. His mean RT for the .5-sec. interval was 94 msec. greater than for the 1.0-sec. interval. Telford's results seemed so far out of line with the other tests that we repeated his study using eight trained *S*s and about five times as many total RT's. In our results the .5-sec. interval gave the lowest RT but it was only 7 msec. below the 1.0-sec. interval mean RT. Our results were in complete agreement with the earlier statement that individual foreperiods have little effect in tests in which foreperiods are randomly mixed.

Check tests.—A different type of foreperiod variability from any discussed above is frequently used in RT studies and should be mentioned. It is the use of a small number of infinite foreperiod (check) tests in a series of constant or nearly constant foreperiods. These infinite foreperiods produce no RT's and are usually neglected in any calculation of foreperiod variability and mean. A more reasonable procedure might be to use a small number of very long foreperiods for check tests which would allow some calculation of foreperiod variability or to treat the infinite foreperiods as separate no-stimulus—no-response alternatives in a choice reaction test.

Summary

Six *S*s were given two series of simple reaction-time tests. In the first series the effect of changes in mean foreperiod and foreperiod variability were systemmatically investigated. In the second series the effect of spacing between stimuli was studied with no warning signal. These tests were designed to determine the relation between RT and *S*s' uncertainty about time of stimulus presentation.

The results show that RT increases with foreperiod variability and with mean foreperiod above some small optimum value less than 1 sec. In a sequence of trials, the immediate foreperiod influences RT only if the previous foreperiod is different from it, and then only slightly.

The striking finding in all tests with variable foreperiod is that the important determiner of RT is not the immediate foreperiod but rather the distribution of foreperiods within which it is embedded.

References

1. Breitwieser, J. V. Attention and movement in reaction time. *Arch. Psychol., N. Y.*, 1911, **18,** 18–38.
2. Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4,** 11–26.
3. Teichner, W. H. Recent studies in simple reaction time. *Psychol. Bull.*, 1954, **51,** 128–149.
4. Telford, C. W. The refractory phase of voluntary and associated responses. *J. exp. Psychol.*, 1931, **14,** 1–36.
5. Woodrow, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17,** No. 5 (Whole No. 76).
6. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 29, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

CHANGES IN RESPONSE STRENGTH WITH CHANGES IN THE AMOUNT OF REINFORCEMENT

ROBERT H. DUFORT AND GREGORY A. KIMBLE

Duke University

In *Principles of Behavior* (**5**), Hull attributes decrement in response strength during experimental extinction to the accumulation of reactive inhibition (I_R) and conditioned inhibition ($_SI_R$). Reactive inhibition is a temporary negative motivational state resulting from response evocation; conditioned inhibition is a permanent habit of not responding, reinforced by I_R-reduction during intervals when no response occurs. Both I_R and $_SI_R$ subtract from reaction potential ($_SE_R$) and response decrement results. Extinction is not weakening of habit strength ($_SH_R$) but a performance decrease.

Within the Hullian system, there are several alternatives which might be used to account for extinction in terms of a reduction in $_SH_R$. Of these, the most obvious, as pointed out earlier by Zeaman (**11**), would consider extinction to be the result of decreasing the amount of reinforcement to zero. Although Hull never specifically considers the effect upon $_SH_R$ of decreasing the amount of reinforcement, he does treat the case of increasing amounts. His interpretation there is that the effect of increasing the amount of reinforcement (w) is to change the limit (M) of habit growth. Moreover, he presents an equation for describing this change which can presumably be generalized to cover the case of amount decrease as well. The equation is:

$$_SH_R = (M' - {}_SH_{R'})(1 - \epsilon^{-iN}) + {}_SH_{R'} \quad (1)$$

where $M' = M(1 - \epsilon^{-kw})$

and
- N = number of reinforcements
- i, k = constants
- w = amount of reinforcement
- ϵ = 10
- M = the maximum of habit strength, taken as 100
- M' = the maximum of habit strength modified by the amount of reinforcement
- $_SH_R$ = habit strength at any given reinforcement after the amount change
- $_SH_{R'}$ = habit strength developed at the time of the amount change.

The nature of this equation is such that changing the amount of reinforcement alters the limit approached by $_SH_R$. Whether the change is an increase or decrease, $_SH_R$ approaches the new maximum (M') as a growth function of N. If the amount of reinforcement is shifted to zero, the limit of the decreasing function is also zero. The extinction curve thus becomes the limiting case of a family of curves obtained by decreasing the amount of reinforcement. Moreover, Hull's theory demands that the slopes of these functions be the same. That for the extinction curve should be identical with the slopes of other functions obtained with increases and decreases in the amount of reinforcement.

Several studies (**2, 3, 4, 10**) have demonstrated differences in response strength produced by differences in the amount of reinforcement. Others (**1, 11**) have shown that response strength increases or decreases appropriately with increases or decreases in amount. These studies have stressed certain emotional (frustration and elation) effects occurring with

changes in amount. None has presented a detailed analysis or comparison of curve forms obtained after the change. Nor has any obtained data relevant to the question whether the extinction process can be attributed to a decrease in the amount of reinforcement to zero.

The purpose of this study is to test the prediction that the data obtained for both increases and decreases in amount can be satisfactorily represented by the same curve forms, and that the curve obtained during experimental extinction will differ from such functions only in asymptote.

Method

Subjects.—The *S*s were 40 experimentally naive male albino rats of the Wistar strain, ranging in age from 100 to 180 days. They were received from the supplier in two groups of 16 and 24, the second group arriving 15 days after the first group had completed the experiment. Upon receipt they were housed in individual living cages and allowed free access to food and water for 4 days. During this time each *S* was handled for 10 min. a day. On Day 5 the *S*s were placed on a 23-hr. feeding schedule, having access to food 1 hr. a day at the time their squad would finish the daily experimental sessions. Each *S* continued to have free access to water. Experimentation began on the Day 8.

Twenty *S*s ran in a morning (A.M.) squad and 20 in an afternoon (P.M.) squad. Assignment to these squads was at random. In the first group, 6 *S*s ran in the A.M. squad and 10 in the P.M. squad; in the second group, 14 ran in the A.M. squad and 10 in the P.M. squad.

Apparatus.—The apparatus was an elevated stand, consisting of a 10 × 5-in. starting platform, an 18 × 2.5-in. straight runway, and an approximately semicircular goal platform 16 in. long and 30 in. wide. The straight runway led into the goal platform at the center of its straight side. The starting platform was separated from the runway by a removable block which prevented *S* from leaving the starting area until he met an orientation criterion.

Five soft drink bottle caps were arranged at equal intervals around the circumference of the goal platform, at a distance of 15 in. from the end of the runway. The center cap was directly in line with the center of the runway. The others were located at 5.75 in. and 11.50 in. on either side of the center cap. Each cap had a slight dent in the top to hold the fluid reinforcement. The goal cap immediately to the right of the center cap was the correct one.

A black cardboard 17 in. high, supported by hardware cloth, was attached to the curved edge of the goal platform. The cardboard prevented *S* from jumping off the apparatus and provided a homogeneous black background, except for a 4 × 5-in. white card, which was directly behind the correct cap. This card remained in place throughout the experiment, except for pretraining, in the belief that it would provide a distinctive cue and facilitate learning. The reinforcement used in the experiment was one drop (.08 ml.) of sucrose solution. To obtain different amounts of reinforcement, the solutions were of 5%, 10%, and 20% concentration. They were prepared by adding to a given amount (e.g., 5 mg.) of sucrose enough distilled water (e.g., 95 ml.) to produce 100 gm. of solution. As Guttman (4) has pointed out, the advantage of using nutrients in solution is that this avoids confounding the amount of reinforcement with the amount of consummatory activity.

The apparatus provided a measure of the time required for *S* to leave the starting platform and run down the straight alley to the goal platform, a distance of 18 in. Removing the block separating the starting platform from the runway closed a switch which started a .01-sec. timer. The timer stopped when *S* broke a beam of light directed at the element of a photocell, located at the end of the runway. In order to reduce the visibility of the beam it was passed through a bottle of household iodine.

Procedure.—The experimental procedure involved three stages: pretraining, original learning in which all *S*s received the same amount of reinforcement and, finally, a test condition in which the amount of reinforcement varied for different groups.

Pretraining.—The experiment began with a series of pretraining sessions. The *E* placed *S* on the starting platform and waited for it to orient toward the goal platform. When this happened, *E* removed the block and allowed *S* to explore the apparatus for 10 min. On these trials each goal cap contained one drop of 10% sucrose solution. The *E* recorded running time, cap preference, and noted whether *S* drank the solution in each cap. This procedure was followed daily until *S* met two criteria on two successive days: (*a*) a running time of 20 sec. or less and (*b*) drinking from all five caps.

Original learning.—In this stage of the study, *S* learned to run to the correct goal cap, which contained one drop of 10% sucrose solution; the other caps were empty. The *S* received 10 trials a day for two successive days. In each daily

session, trials were separated by about 15 sec. The procedure followed that used in pretraining. The *E* removed the block separating the starting platform from the runway after *S* oriented toward the goal platform and recorded running time and the goal cap to which *S* responded. In this latter case *E* counted a response if *S* actually touched a cap as if to drink; contacts with the feet and tail did not qualify as responses. If *S* responded to the wrong cap, he was returned to the starting platform for another trial; correction was not permitted.

Experimental treatment.—At the completion of the original learning sessions, *Ss* were subdivided into four groups of 10 *Ss* each and given four daily sessions of 10 trials each with different amounts of reinforcement. One group continued on the 10% solution. The others were switched to 20%, 5%, and 0% conditions. The *Ss* in the 0% (extinction) condition found the goal cap empty. The *Ss* were assigned to these conditions prior to original learning in a counterbalanced order depending upon when they reached the criteria terminating pretraining.

From preliminary experimentation, it was suspected that in the 0% group *Ss* would cease to run near the end of the 40 trials. In order to prevent interference with the experimental schedule, a criterion of 20 sec. for terminating a trial was imposed on all groups. It was necessary to invoke this criterion only on *Ss* in the 0% group. This meant that there was an artificial upper limit of 20 sec. on the latencies of these *Ss*.

Results

The basic data from the experiment appear in Fig. 1 and 2, in the form of learning curves. In each case the trials are grouped in blocks of four to reduce irregularity in the functions. In spite of this, the probability-of-response data in Fig. 1 are too uneven to justify precise mathematical treatment. The important points concerning these data seem to be as follows: (*a*) There is a clear separation of the four functions on this measure. An analysis-of-variance testing for over-all differences in the functions obtained after the amount change produced an F of 34.34, ($P = .001$). (*b*) Although the 5% sucrose solution is well above the reinforcement threshold estimated by Guttman, there is a decrease in percentage of correct

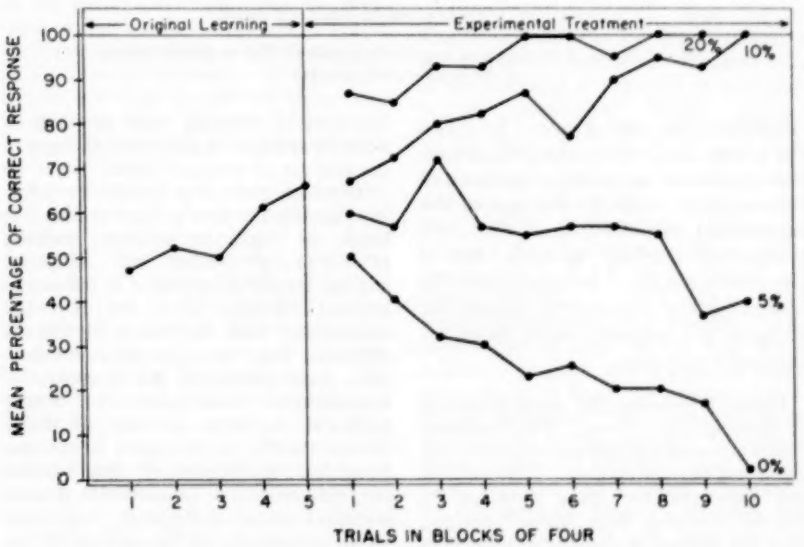

Fig. 1. Mean percentage of correct response as a function of practice. Original learning data are for 40 *Ss*. For each experimental treatment, $N = 10$.

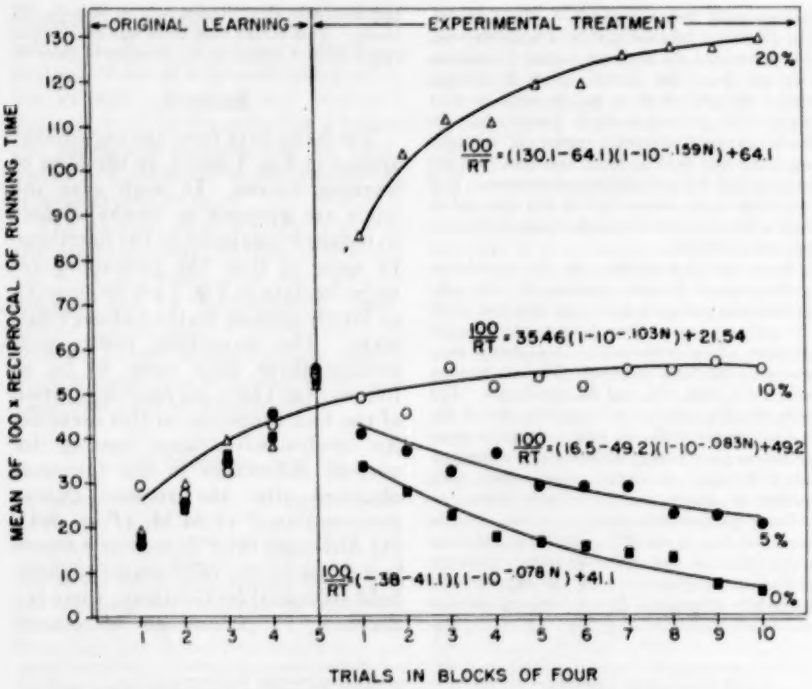

FIG. 2. Speed of running as a function of practice. For an interpretation of the fitted curves, see text.

responses for this group. (*c*) The shift from the 10% to the 20% condition produced an abrupt increase in this measure. (*d*) At the end of the experiment the 10% and the 20% groups had reached the same level of response strength. In this repect the results for the probability-of-response measure are different from those for the speed measure.

Figure 2 presents the speed measures in learning curve form. The empirical points are means of 100 multiplied by the reciprocal of running time. The vertical line separates the data obtained in original learning from those obtained after the change in amount of reinforcement. Renumbering of the blocks of trials after the change is to accommodate the type of function used by Hull to describe changes in ${}_{S}H_{R}$ with changes in the amount of reinforcement.

Although there is a significant difference among the four groups on the first block of trials in original learning ($F = 4.13$; $P = .025$), this difference has completely disappeared at the end of original learning ($F = .39$). On the assumption that the initial significant difference was a sampling error, the data have been combined for purposes of mathematical treatment. In fitting curves to the data, we have fit Hull's curves directly to the speed of running measures. As Spence (9) has pointed out, this procedure is justifiable if such measures show differences over and above differences in latency or if the learning curves do not reflect the early stages of learning. The length of the

runway used in the present study, and the fact that the percentage of correct responses was at 50% after only four trials would seem to justify these assumptions. Another consideration which cannot be overlooked is that growth functions fit the data reasonably well.

One of the implications of Hull's theory is that the slopes of the functions obtained with a shift in the amount of reinforcement should be the same as the slope of habit growth obtained without a shift. For this reason, the curve fit to the data for the 10% control group is of the form, ${}_{S}H_{R} = M(1 - e^{-iN})$, used by Hull to depict habit growth with no change in reinforcement amount. It covers the entire series of 15 blocks of trials. The curves fit to the other three sets of data, obtained after the change in amount of reinforcement, are of the form of Equation 1. The curve-fitting procedure was that described by Lewis (7).

With the data in this form, it is possible to evaluate the adequacy of Hull's theory for handling them. First of all it may be pointed out that the growth curves fit the data rather well. For the functions in the right hand portion of Fig. 2, an F test for goodness of fit reveals that in no case is it possible to reject the hypothesis that the growth curve adequately represents the data. The largest F was 1.265 which is not significant even at the .20 level.

According to Hull's theory, changing the amount of reinforcement should produce new asymptotes of habit strength. The test of this prediction proceeded in a series of steps. To determine whether the functions had reached asymptotic values, a treatments-by-Ss analysis of variance was applied to the last three points on each curve. The largest F produced by this analysis was 2.359 for the 0% group. This F is significant at only the .20 level. All other F values were less than 1.0. A second step in this analysis involved t tests to determine whether the empirical asymptotes are different from the asymptotes estimated in the curve-fitting procedure. Comparison of these values for each group showed the t of 114.6 obtained for the 0% group to be highly significant. Presumably this was the result of using a criterion latency in the case of this group. The other t values were all less than 1.0. To determine whether the empirical asymptotes are significantly different, an analysis of variance involving the last 3 blocks of trials for all groups was employed. The F of 89.49 is significant at the .001 level. To determine whether adjacent asymptotes are significantly different, t tests were employed on the same measures. The t values obtained were: for a comparison of the 0% group and 5% group, 3.40; for the 5% group and the 10% group, 3.29; for the 10% group and the 20% group, 7.33. These t values are all significant at the .01 level or better. On the basis of this series of tests, the conclusion that groups with different sucrose concentrations reached different asymptotes seems entirely justified.

The remaining question is whether the slopes of the four functions are the same. It was predicted that data obtained for both increases and decreases in amount could be represented by Equation 1. The critical value in the fitted equations for testing this prediction becomes that of the constant i, the rate of approach to the asymptote. Within Hullian theory, a change in the amount of reinforcement changes the asymptote of the learning curve, but does not change the rate of approach constant. If the curves are identical, the values of i should be similar. As a test of this, the i values for the 0% and 5% curves, which are almost identical (−.078 and −.083), were averaged and the resulting value substituted in each of the four fitted equations in place of the i values which were obtained from the original curve-fitting procedure. The goodness of fit test applied to the resulting calculated values indicated that the growth functions with the new i values adequately fit the empirical data in all cases but 20%. In this case, the hypothesis of adequate fit must be rejected with an F significant at the .001 level ($F = 5.189$); the new calculated curve underestimated

the original values at every point on the function.

Discussion

The curve-fitting analysis of the results of this experiment indicates that deductions from Hull's theory describe the data very well except for those obtained from the group shifted from a 10% to a 20% solution. In this case the change in behavior strength was much greater and more rapid than would be predicted either from the function obtained early in learning or from the rest of the data obtained after the shift in amount of reinforcement. The general suggestion of the data is that the effects produced by increasing the amount of reinforcement are different from those produced by decreasing the values of this variable. It may be that the rapid increase in speed of responding for the group receiving an increased amount is an exaggerated elation effect of the sort described by Crespi (1). Since there was no 20% control group in this experiment, we cannot say. If the explanation for this feature of the results does ultimately turn out to be of this sort, the indication is that the effect of changing the amount of reinforcement is at least partly on performance rather than habit. Such an account would be more nearly in line with Hull's (6) recent treatment in terms of incentive motivation (K) than it would be with the theory examined in this paper which treats the results in terms of changing asymptotes of ${}_{S}H_{R}$.

At a more general level the important point of this paper stands. Whether the effect of changing the amount of reinforcement is on habit or performance, it seems quite clear that the effect must, almost of logical necessity, cover a part of the explanation of extinction. There is now ample evidence that increasing and decreasing the amount of reinforcement increases and decreases response strength appropriately. It seems a theoretical extravagance to assume a completely new process to deal with the special case of decreasing the amount to zero. Furthermore, the fact that the slope of the extinction function obtained in this experiment is the same as that for the curve produced by decreasing the amount to 5% indicates in a factual way the fruitfulness of considering extinction as a part of the same process as is involved in decreasing amounts of reinforcement.

In Hull's theory there are two other ways to reach an explanation of extinction similar to that developed in this paper. The alternative accounts would view extinction as the result of increasing the CS-UCS interval or the delay of reinforcement to a point where no learning occurs. Hull did not treat the effect upon behavior of changes in these variables. It was largely for this reason that this experiment manipulated the amount of reinforcement. At the same time, it seems likely that changes in either variable would produce changes in performance, and that a large increase could produce experimental extinction. In the case of the CS-UCS interval, there is experimental evidence which points directly toward such an interpretation. McAllister (8) found that increasing the interstimulus interval from an optimal to a nonoptimal value produced a decline in performance resembling extinction.

Summary

The purpose of this study was to determine whether, as one aspect of Hull's theory would indicate, the slopes of performance curves produced by changes in the amount of reinforcement would be the same whether the change was an increase, decrease or even a decrease to zero (experimental extinction).

The *S*s in the experiment were 40 albino rats. They were trained to run down an 18-in. alley to obtain a 10% sucrose solution at one of five goal cups. After 20 such trials *S*s were divided into four experimental groups which received a 5%, 10%, or 20% solution or nothing.

The different groups behaved differently during this phase of the experiment on both of the measures available in the experiment. Because of an unsatisfactory degree of irregularity in the percentage of response measure, however, the analysis is largely confined to the speed of running data. With one exception these data fit the predictions drawn from Hull's theory very well. As the theory predicts, shifting the

amount of reinforcement alters the asymptotes approached by the learning curves. Moreover, the functions change at the same rate except for the group in which the shift was from 10% to 20% sucrose. In this case the increase in performance was much greater than the theory would predict. This finding suggests that the influence may be on performance rather than on learning as assumed in this paper. It does not, however, alter the major conclusion that extinction may be treated as the limiting case of a reduction in the amount of reinforcement.

References

1. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55,** 467–517.
2. Fletcher, F. M. Effects of quantitative variation of food-incentive on the performance of physical work by chimpanzees. *Comp. Psychol. Monogr.*, 1940, **16,** No. 82.
3. Grindley, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, **20,** 173–180.
4. Guttman, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.*, 1953, **46,** 213–224.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
6. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univ. Press, 1951.
7. Lewis, D. *Quantitative methods in psychology.* Iowa City: Gordon Bookshop, 1948.
8. McAllister, W. R. The effect on eyelid conditioning of shifting the CS-US interval. *J. exp. Psychol.*, 1953, **45,** 423–428.
9. Spence, K. W. The relation of response latency and speed to the intervening variables and *N* in S-R theory. *Psychol. Rev.*, 1954, **61,** 209–216.
10. Wolfe, J. B., & Kaplon, M. D. Effect of amount of reward and consummatory activity on learning in chickens. *J. comp. Psychol.*, 1941, **31,** 353–362.
11. Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39,** 466–483.

(Received April 8, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

CONDITIONED STIMULUS INTENSITY AND TEMPORAL FACTORS IN SPACED-TRIAL CLASSICAL CONDITIONING [1]

GERALD W. BARNES [2]

Indiana University

The effects of varying temporal relationships between a conditioned stimulus (CS) and an unconditioned stimulus (UCS) on some measurable aspects of behavior have received intensive treatment in the theorizing of Hull (**12**) and Guthrie (**9**) and in much experimental work (**10, 21**). In general, one may distinguish two types of these relationships which have been examined: (*a*) those in which CS onset precedes UCS onset and CS termination precedes or coincides with UCS termination, and (*b*) those in which UCS onset precedes CS onset. Very little evidence is available, however, concerning a third type of CS-UCS relationship: that in which CS onset precedes UCS onset but CS termination occurs after, i.e., "overlaps," UCS termination.

Contiguity theory (**9**), stressing the role in conditioning of stimulation changes, apparently leads to a straightforward prediction of differences in "strength of conditioning" as some inverse function of CS-UCS overlap duration. To the extent that a once-neutral stimulus has acquired the properties of a reinforcing stimulus, the CS-UCS overlap relationship would seem to appear in Hullian (**14**) and Hull-derived (**5**) theories largely as a delay-of-reinforcement variable. The only direct investigation of the CS-UCS overlap relationship (**19**) suggested an inverse relation between performance level and length of CS-UCS overlap, but its use of an operant response vitiates these data appreciably. The present study is partially intended to clarify the nature of this relationship.

Another phenomenon of some interest is the effect of CS intensity on conditioned responding. Hull (**13**) discussed this variable under the label "stimulus intensity dynamism," a concept later incorporated into his theory (**14**) as a postulate which states, briefly, that reaction potential is a direct function of CS intensity. The literature on this subject presents a by-no-means consistent picture. Some *Es* report a direct relationship between CS intensity and level of responding (**13, 15, 16, 22**), others report no differences or mixed effects (**2, 4, 6, 7, 8, 17, 23**), while still others report an inverse relationship for some intensity values (**1, 18**). The second major purpose of the present experiment is examination of the influence on behavior strength of widely disparate CS intensities.

METHOD

Subjects and apparatus.—The *Ss* were 40 dogs unselected for age, size, weight, sex, or breed. The *Ss* were housed in cages in a well-heated and well-lighted room adjoining the experimental room. Water and Purina Kibbled Meal or Dog Checkers were available in the cages at all times.

[1] This article is based on parts of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. Degree in the Department of Psychology, Indiana University. The author is greatly indebted to Professors C. J. Burke and D. W. Lauer for valuable assistance during the course of the study. Thanks are also due the author's colleagues Professors J. J. Antonitis and A. D. Glanville for critically reading the manuscript.

[2] Now at the University of Maine.

A restraining stock, placed in the 8 × 8 × 72/3 ft. high semi-soundproof experimental chamber, was used to partially immobilize *S*. At the front of the stock was an adjustable pillory-like device through which *S*'s head could be inserted and retained. Foreleg movement was restricted by fastening an anchored web-strap around the ankle of each foreleg. A broad canvas band passed under the abdomen and a narrow canvas strap, attached to the rear of the broad band, passed between the hind legs prevented *S* from adopting a sitting or crouching position in the stock.

A 60-cps sine-wave electric shock could be delivered to *S*'s right hind paw through two Grass EEG electrodes. The shocking device employed was set to deliver 1.7 ma. through a 70,000-ohm dummy resistance; some slight variation in current output was present as a function of resistance changes between poles.

The CS used was an 800-cps tone produced by a Hewlett-Packard Model 200b audio-oscillator. The oscillator output was delivered to a 6-in. Jensen speaker mounted on top of the stock. CS intensity was manipulated by adjusting the gain control of the oscillator. Gain-control settings were located giving intensities of 60 db and 80 db *re* .0002 dynes/cm.^{2} at about the level in the stock of the head of a medium-sized *S*.

Onset, duration, and termination of the CS and UCS were determined by an automatic electromechanical arrangement. This provided CS durations of 1, 6, 16 ± .10, and 31 ± .05 sec. and presentation of a .1-sec. shock at a .9-sec. interval after any CS onset. (All temporal intervals were measured with .01-sec. Standard Electric clocks.) By this means, CS-UCS overlaps of 0, 5, 15 ± .10, and 30 ± .05 sec. were defined. CS and UCS onset, duration, and termination were recorded on one channel of a Brush direct ink-writing multichannel oscillograph with a chart speed of 5 mm./sec. The relays involved in the presentation and recording circuits were mounted in a semi-soundproof box located outside the experimental chamber.

The response measured was the angle which the right femur made with the body of *S*. A saddle-like device, fitted onto *S*s back, held a linear wire-wound potentiometer. Half of an ordinary hacksaw blade was attached by a set screw at the pivot point of the control shaft of the potentiometer. When the set screw was centered over *S*s right lesser trochanter (by palpation of the joint) and the hacksaw blade taped to the upper hind leg so as to follow the long axis of the femur, angular movement of the femur was translated into a change in a small DC voltage being impressed on the potentiometer. This voltage change was amplified and led into one channel of the Brush oscillograph, thus giving a record of angular rotation of the femur.

Procedure.—It was originally planned to run *S*s in four successive squads of 10 *S*s each. A fifth squad was made necessary, however, by the loss of 8 *S*s. Five *S*s were assigned at random to each of the eight cells of a 4 × 2 factorial design. Members of the replacement squad were assigned at random to the positions left vacant by lost *S*s.

Each *S* was given one habituation trial a day for the first four days. As soon as *S* had been restrained in the stock and the recording equipment attached, hair was clipped from a small area on the back of the right hind paw extending approximately 2 in. upward from the rear foot pad. Two Grass EEG electrodes were fixed ½ in. apart on a strip of adhesive, filled with electrode paste, and taped to the clipped area. The *E* then left the experimental room and allowed *S* to remain in the stock for 10–12 min.

Days 5–18 constituted the experimental days. The first *S* to be run on a given day was harnessed in the stock, the rear of the right hind paw clipped, and all apparatus attached to him. The *E* then left the experimental room, selected the appropriate CS-UCS overlap and CS intensity, started the recording equipment, and gave *S* a single paired CS-UCS presentation. Twenty to 30 sec. following CS termination, *E* re-entered the experimental room, removed the *S*, and returned him to his home cage. The second *S* to be run was then brought into the experimental room, and the same procedure followed. This method was continued until all members of the squad had received a single experimental trial. This procedure was then repeated with the same running order until each *S* had received either four trials (on experimental Days 1–7) or three trials (on experimental Days 8–14). By this method, 49 experimental trials were given each *S* with a minimum intertrial interval of 2 hr.

The same general procedure was followed for each squad. Running orders were determined at random for each day, but they were cyclical within a given day.

Results

A "response" was defined as any femoral rotation of 5° or more from base level, where base level was determined from performance during the 3 sec. preceding onset of the CS. The base level, then, was used as an arbitrary zero point from which femoral rotations were measured. Any response which began after CS onset but before UCS onset was defined as a CR. Thus, occurrence and latency (to nearest .1 sec.) of CR's were recover-

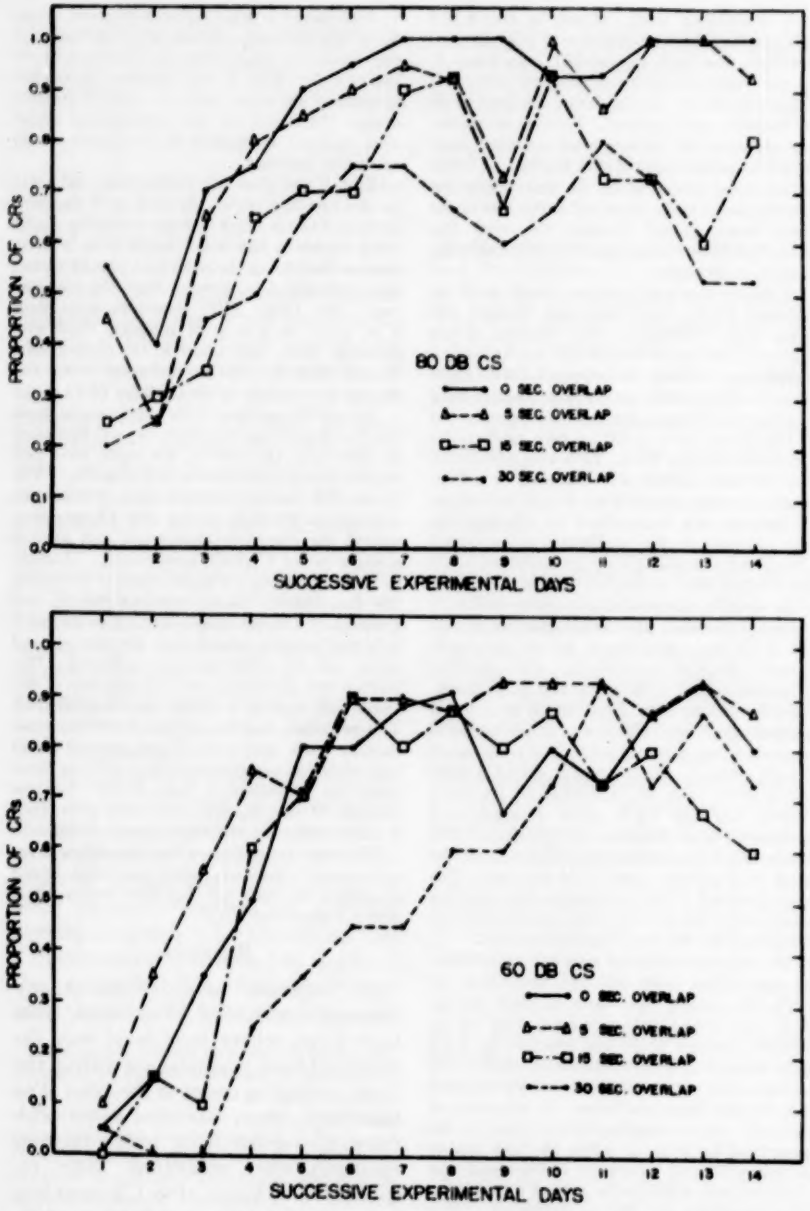

FIG. 1. Daily proportions of CR's made by each group. Points for Days 1–7 and for Days 8–14 represent proportions of CR's for five *Ss* over four and three daily trials, respectively.

able from the records for each experimental trial.[3]

The proportions of CR's made by the 80-db and 60-db CS groups on each experimental day are presented in the upper and lower panels of Fig. 1. Ordering of the 80-db groups is reasonably clear, the longer overlap groups showing consistently smaller proportions of CR's than the shorter overlap groups.

The 60-db groups do not, however, present such a straightforward picture. The curve for the 5-sec. overlap group is above the remaining three group curves with some consistency, but inspection of the other 60-db groups indicates little difference between the 0-sec. and the 15-sec. overlap groups until possibly Day 12. The 30-sec. overlap group is well below all the other groups in proportion of CR's until Day 11. It is of some interest to note that the 30-sec. overlap group reaches a level of responding as high as (or higher than) the 15-sec. overlap group. Of further interest is the observation that, within each CS intensity, the slopes of the group curves appear very similar.

Comparison of the upper and lower sections of Fig. 1 indicates that the two shortest CS-UCS overlap groups receiving the 80-db CS gave more CR's than their corresponding 60-db CS groups throughout the experiment. The two 15-sec. overlap groups, however, are remarkably similar after Day 5 and, in the 30-sec. overlap cases, the 60-db group even appears to have surpassed its corresponding 80-db group in proportion of CR's after Day 9.

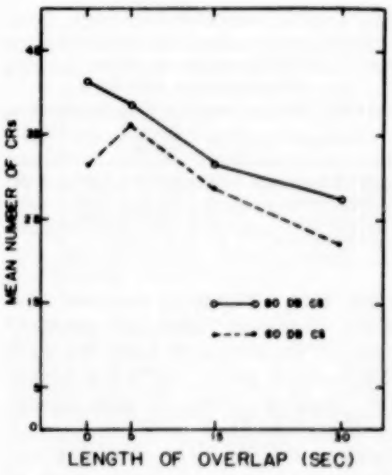

FIG. 2. Mean total number of CR's made by each CS intensity group as a function of duration of CS-UCS overlap.

Figure 2 presents the mean total numbers of CR's made by each group in all experimental trials. The ordering of the groups in terms of both CS-UCS overlap and CS intensity is striking, with each 80-db group above its corresponding 60-db group and, with a single exception in the 60-db groups, smaller numbers of CR's shown as CS-UCS overlap increases within CS intensity. As previously indicated, however, the order in terms of CS intensity may not reflect uniformly differences maintained over the entire course of conditioning. The results of an analysis of variance performed over these data are shown in Table 1, this analysis yielding F values significant beyond the .05 level for both CS-UCS overlap and CS intensity.

Trial-by-trial latency data were so similar to the over-all picture presented by proportions of CR's that it does not seem necessary to include

[3] The laboratory in which the experiment was performed was flooded, resulting in loss of the records of one trial for each of six *Ss*. In these cases, CR probabilities for *Ss* concerned were estimated from the three trials preceding and the three trials following the lost trial, and a random number was drawn to define CR occurrence or nonoccurrence. If a CR was indicated by this Monte Carlo run, the median latency of the three trials preceding and following the lost trial was taken to be the latency for that trial.

TABLE 1

ANALYSIS OF VARIANCE OF MEAN TOTAL CR's MADE BY THE EIGHT EXPERIMENTAL GROUPS

Source	*df*	Mean Square	*F*	*P*
CS-UCS overlap	3	346.97	7.70	<.01
CS intensity	1	270.40	6.00	<.05
Interaction	3	30.60	0.68	
Within groups	32	45.08		

TABLE 2

ANALYSIS OF VARIANCE OF MEAN TOTAL RECIPROCAL LATENCIES EXHIBITED BY THE EIGHT EXPERIMENTAL GROUPS

Source	*df*	Mean Square	*F*	*P*
CS-UCS overlap	3	6.30	4.56	<.01
CS intensity	1	7.44	5.38	<.05
Interaction	3	1.28	0.93	
Within groups	32	1.38		

them here. Figure 3, however, presents reciprocal latencies averaged over all experimental trials for each experimental group, and it is seen that the resulting curves are quite similar to those obtained with mean total number of CR's. (In obtaining raw latency values, an arbitrary value of infinity was assigned for trials on which a CR did not occur.) Table 2 presents the analysis of variance of these "mean total reciprocal latencies," this analysis yielding *F* values significant beyond the .05 level for both major variables.

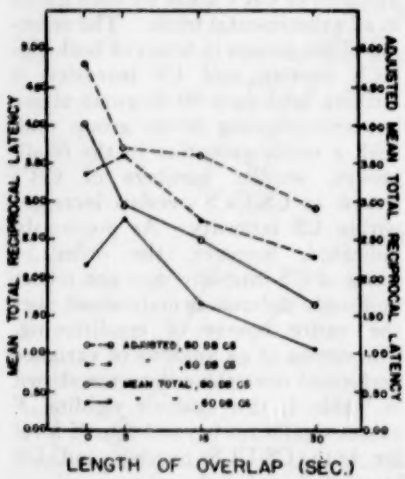

FIG. 3. Mean total reciprocal latencies and adjusted mean total reciprocal latencies shown by each CS intensity group as a function of duration of CS-UCS overlap. These quantities are defined in the text.

It has been noted elsewhere (**3**) that response probability and response latency are, in general, related measures. Since in the present experiment a conditioned femoral rotation of 5° or more was required to occur within .9 sec. after CS onset, the latency distribution of this response is necessarily truncated at that point. A latency measure which appears relatively independent of CR probability results from summation of the reciprocal latencies for a given *S* over the 49 trials and division of this sum by the total number of CR's made by that *S*. Such numbers, which reflect differences in latency for trials on which Ss gave CR's, will hereafter be called "adjusted mean total reciprocal latencies." The curves obtained from such measures are also presented in Fig. 3, and we see that the curve forms are only slightly different from those for mean total reciprocal latencies alone. The analysis of variance of these data is shown in Table 3, indicating that differences due to CS-

TABLE 3

ANALYSIS OF VARIANCE OF ADJUSTED MEAN TOTAL RECIPROCAL LATENCIES EXHIBITED BY THE EIGHT EXPERIMENTAL GROUPS

Source	*df*	Mean Square	*F*	*P*
CS-UCS overlap	3	4.01	2.94	<.05
CS intensity	1	5.59	4.10*	
Interaction	3	0.74	0.54	
Within groups	32	1.36		

* $F = 4.15$ required for significance at the .05 level.

UCS overlap are again significant beyond the .05 level, and that differences due to CS intensity fall just short of significance at this level.

Discussion

The obtained effects of CS intensity are not in great conflict with the stimulus intensity dynamism notion proposed by Hull (**13**). When total number of CR's, mean total reciprocal latencies, and adjusted mean total reciprocal latencies are considered, a striking dynamism effect is clearly seen. Possible deviations were found, however, in that the final levels of responding reached by the corresponding long overlap groups were quite similar. If these results are confirmed by further experimentation, some modification of the dynamism postulate is indicated.

The conclusion is clearly warranted that CS-UCS overlap is a variable affecting behavior strength. This result seems relevant for several contemporary theories of learning. For example, contiguity theory holds that "a combination of stimuli which has accompanied a movement will on its recurrence tend to be followed by that movement" (**9**, p. 26). At a qualitative level, one would expect on this basis that the number of CR's performed under the conditions earlier described would be some rapidly decreasing function of duration of CS-UCS overlap. It would be of some importance to know whether or not *S*s in the nonzero overlap groups continued to respond during CS presentation after the first response—whether CR or UCR—had been completed. Since the records appear to show that *S*s were not, on the average, responding at CS termination, contiguity theory would apparently predict that they should learn not to respond in the presence of the CS. Hence, conditioning should become progressively poorer as duration of CS-UCS overlap increases.

On closer analysis, however, a major difficulty for a contiguity interpretation seems posed by the relatively high final levels of responding reached by the two longest overlap groups (Fig. 1), which it would seem should exhibit little if any evidences of conditioning on contiguity grounds. One possible explanation may lie in use of a variant of the fractional-goal-response concept introduced by Hull (**11**). If the contiguity theorist postulates that fractional components of the leg-flexion response were occurring with diminishing frequency (and/or strength) after the initial gross leg flexion, then the reported results could be qualitatively predicted.

At least one Hullian-type theory (**5**) seems to yield approximately the same prediction as a contiguity theory, although by a more devious route. From the point of view presented by Dollard and Miller (**5**), several opposing variables apparently act in this situation to determine performance level. First, *S*s should, through simple classical conditioning, develop comparable tendencies for paw withdrawal when the CS is presented. *Pari passu*, the CS should acquire anxiety-producing properties by contiguous association with the noxious UCS. The amount of anxiety generated by the CS should, however, be inversely related to length of CS-UCS overlap, for more opportunity is allowed for extinction of the fear response to the CS as overlap length increases. It is assumed that anxiety reduction is reinforcing, and, presumably, we may here identify anxiety reduction with CS termination. Hence, such reinforcement for paw withdrawal is influenced both by its delay and by its differential amount. When we further note that nonresponding in the presence of the CS receives, on the average, immediate reinforcement through anxiety reduction in the nonzero overlap groups, the conclusion is reached that performance level should be an inverse function of length of CS-UCS overlap, but that this level need not be drastically reduced in the case of the overlap lengths herein employed.

This discussion should not be taken as an attempt to establish the present experiment as an *experimentum crucis* for contiguity theory vs. Hullian theories. Such a crucial experiment, indeed, seems impossible at the level of these theories, for the propositions contained in them

are not formulated so as to allow an investigator to apply them to specific experimental situations. The discussion does indicate, perhaps, that two popular learning theories might be sharpened somewhat.

Summary

By use of a classical conditioning situation in which a brief 1.7-ma. shock applied to the right hind paw of dogs was paired with an 800-cps pure tone, the relationship of frequency and latency of conditioned leg-flexion responses to (*a*) two widely separated CS intensities and (*b*) four values of the time by which CS termination extended beyond ("overlapped") shock termination was examined. During the experiment proper four trials a day were given each *S* on Days 1–7 and three trials a day were given on Days 8–14, with a minimum intertrial interval of 2 hr.

Mean total number of CR's and mean total reciprocal latency measures were found to vary directly with CS intensity and to vary inversely, with a single inversion of order, with length of CS-UCS overlap. Some anomalies, however, were noted in consideration of day-by-day behavior.

Implications of the results for Hullian and contiguity theory were discussed.

References

1. Baker, L. E. The pupillary response conditioned to subliminal auditory stimuli. *Psychol. Monogr.*, 1938, **50,** No. 3 (Whole No. 223).
2. Bragiel, R. M., & Perkins, C. C., Jr. Conditioned stimulus intensity and response speed. *J. exp. Psychol.*, 1954, **47,** 437–441.
3. Burke, C. J. A theory relating momentary reaction potential to response latency. *Psychol. Rev.*, 1949, **56,** 208–223.
4. Carter, L. F. Intensity of conditioned stimulus and rate of conditioning. *J. exp. Psychol.*, 1941, **28,** 481–490.
5. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
6. Fink, J. B., & Patton, R. M. Decrement of a learned drinking response accompanying changes in several stimulus characteristics. *J. comp. physiol. Psychol.*, 1953, **46,** 23–27.
7. Grant, D. A., & Schneider, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38,** 690–696.
8. Grant, D. A., & Schneider, D. E. Intensity of conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39,** 35–40.
9. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
10. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.
11. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, **38,** 487–506.
12. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
13. Hull, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56,** 67–76.
14. Hull, C. L. *A behavior system.* New Haven: Yale Univ. Press, 1952.
15. Kessen, W. Response strength and conditioned stimulus intensity. *J. exp. Psychol.*, 1953, **45,** 82–86.
16. Kupalov, P. S., & Gantt, W. H. The relationship between strength of the conditioned stimulus and the size of the resulting conditioned reflex. *Brain*, 1927, **50,** 44–52.
17. Kupalov, P. S., Lyman, R. S., & Lukov, B. N. The relationship between the intensity of tone stimuli and size of the resulting conditioned reflexes. *Brain*, 1931, **54,** 85–98.
18. Metzner, C. A., & Baker, L. E. The pupillary response conditioned to subliminal auditory stimuli: a control experiment. *Psychol. Bull.*, 1939, **36,** 625. (Abstract)
19. Mowrer, O. H., & Lamoreaux, R. R. Avoidance conditioning and signal duration. *Psychol. Monogr.*, 1942, **54,** No. 5 (Whole No. 247).
20. Perkins, C. C., Jr. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46,** 225–231.
21. Stone, C. P. (Ed.) *Comparative psychology.* New York: Prentice-Hall, 1951.
22. Wilcott, R. C. A search for subthreshold conditioning at four different auditory frequencies. *J. exp. Psychol.*, 1953, **46,** 271–277.
23. Zener, K. Individual differences in the functional relationship of conditioned stimulus intensity and response magnitude. *Amer. Psychologist*, 1949, **4,** 218. (Abstract)

(Received March 31, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

LEARNING AND RETAINING A RATE OF MOVEMENT WITH THE AID OF KINESTHETIC AND VERBAL CUES[1]

ROBERT S. LINCOLN[2]

The Johns Hopkins University

It is generally presumed that kinesthetic sensibility is useful in the skilled regulation of human movements. In tasks involving the movement of controls kinesthetic cues may provide information about the spatial position of the controlling bodily member, its rate of movement, and its acceleration. The value of kinesthetic cues has been assessed in a number of experiments with respect to the accuracy of positioning responses (**1, 4, 7**). Little is known, however, about the effectiveness of kinesthetic information in the maintenance of a rate of movement.

It is possible to train *Ss* to produce a constant rate of rotary movement by supplying them with verbal indications of their rate errors from trial to trial (**5**). Such a procedure forces the trainees to rely heavily upon kinesthetic cues in producing the rate but is contaminated by the various interpretations that the verbal corrections receive.

This experiment was designed to test the ability of human *Ss* to utilize more purely kinesthetic cues in learning to maintain a constant rate of rotary movement. In addition, the retention of the rate was measured as a function of the length of delay between the completion of the learning period and the measurement of retention.

[1] This research was carried out under Contract N5ori-166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-196, Project Designation No. NR 145-089, under that contract. Miss Frances Wolfram aided in the collection of the data.

[2] Now with the RAND Corporation, Santa Monica, California.

METHOD

Experimental Conditions

The experiment consisted of four successive experimental periods: (*a*) the prelearning period, (*b*) the learning period, (*c*) the criterion period, and (*d*) the relearning period.

Prelearning period.—Fifty-four right-handed males were used in the experiment. All *Ss* were told to produce a linear rate of 160 ft. per min. by turning a handwheel in a clockwise direction. Since the handwheel was 15 cm. in diameter, the required linear rate was approximately equivalent to an angular rate of 100 rpm. The *Ss* were given five trials, each 15 sec. long. During these prelearning trials, *Ss* received no knowledge of results.

Learning period.—In the learning period *Ss* received 30, 15 sec. trials during which they continued their attempts to produce the required rate of movement. The learning period was divided into two sessions of 15 trials each, and *Ss* were given a 5-min. rest break at the end of the first session.

For these trials *Ss* were randomly assigned to three groups of 18 *Ss* each which differed in regard to the type of knowledge of results supplied. The *Ss* in one group continued to work with verbal instructions. Before each trial Legan they were told the direction and amount of the average rate error in feet per minute that they had produced during the immediately preceding trial. The second group received kinesthetic error information. These *Ss* grasped the handwheel knob while the handwheel was driven by a motor for 15 sec. at a rate equivalent to their average rate error on the previous trial. The *Ss* were also told whether they had turned too fast or too slowly during the previous trial, but they were not given a verbal indication of the magnitude of their error. The third group of *Ss* did not receive any direct error information. After each practice trial these *Ss* grasped the handwheel knob while the handwheel was driven for 15 sec. at the standard rate of 100 rpm.

The *Ss* in this group presumably inferred their own knowledge of results by comparing their memory of the standard rate with the rate as they produced it in their next trial.

For the two groups that worked with kinesthetic information a delay of 30 sec. occurred between the completion of a practice trial and the start of the information trial. The next practice trial began 15 sec. after the completion of the information trial. Although it was not possible to provide the verbal information group with the same type of intertrial activity, the total time elapsing between learning trials was approximately 65 sec. for all groups.

Criterion period.—Following the learning period all *Ss* attempted to produce the rate for 20 additional trials without further knowledge of results. At this time each training group was subdivided into three smaller groups. One-third of each training group received the criterion trials as soon as the learning period was completed. For convenience this group will be considered as having experienced no delay between the learning and criterion periods. Actually they experienced a 30-sec. delay because of apparatus requirements. A second third of each training group received the criterion trials 1 hr. after completing the learning period. For the remaining third of each training group the criterion trials came approximately 24 hr. after the learning trials. An intertrial interval of 30 sec. was used throughout the criterion period. Within the restrictions necessitated by schedule requirements, *Ss* were randomly assigned to the criterion groups.

Relearning period.—When the criterion period was completed, all groups received five additional trials with the same type of knowledge of results that they had received in the learning trials.

Apparatus and Procedures

The equipment used in this experiment has already been described in a previous report (6). The handwheel, which moved in a vertical plane, was positioned for convenient operation with the right hand. Two measures of performance were obtained. The first was a measure of rate accuracy—the amount of time during a trial that the rate of turning was maintained within ± 9% of the standard rate. The second measure was a count of the actual number of revolutions turned during a trial. An indication of constant error was obtained by subtracting this measure from the standard rate. Since the accuracy scores are an all-or-none measure they do not always correlate with the continuous measure of constant error.

Visual cues of rate were minimized by covering the handwheel with a shield that also covered *S*'s hand and forearm. Auditory cues were minimized by enclosing recording instruments in soundproofed boxes and by taking advantage of a masking noise provided by an air-conditioning system.

A calibrated syncro drive was used to turn the handwheel when kinesthetic information was provided. Because of the limitations of the apparatus, error rates could be duplicated only in steps of two rpm. An equivalent scale unit was used in providing verbal error information. Actually it is most unlikely that *Ss* could utilize all the information even in a scale with this degree of coarseness. A clutch arrangement made it possible to disengage the syncro motor from the handwheel shaft when *Ss* had to turn the handwheel. Friction and inertia in the handwheel were kept at low levels.

The *Ss* knew that the trials would be less than 1 min. long. They were told not to count the number of revolutions while turning the handwheel.

Results

Several analyses of variance were performed on different portions of the data. Differences were accepted as significant if obtained *P* values were equal to, or less than, .05. Error estimates were tested for homogeneity by means of Bartlett's test.

Prelearning period.—The prelearning trials were used to establish a base line of initial rate accuracy. Scores achieved on the fifth prelearning trial were also used to provide knowledge of results for use in the first learning trial. Trial 5 in Fig. 1 is the fifth prelearning trial. The differences among groups on this trial are assumed to be chance differences. These differences cannot reasonably be evaluated by statistical methods because of a preponderance of zero scores. Since *Ss* in all three groups were operating under similar instructions during the prelearning trials, differences between groups would not be anticipated.

During each of the prelearning trials, *Ss* in all groups tended to overestimate the required rate of turning. This overestimation increased during

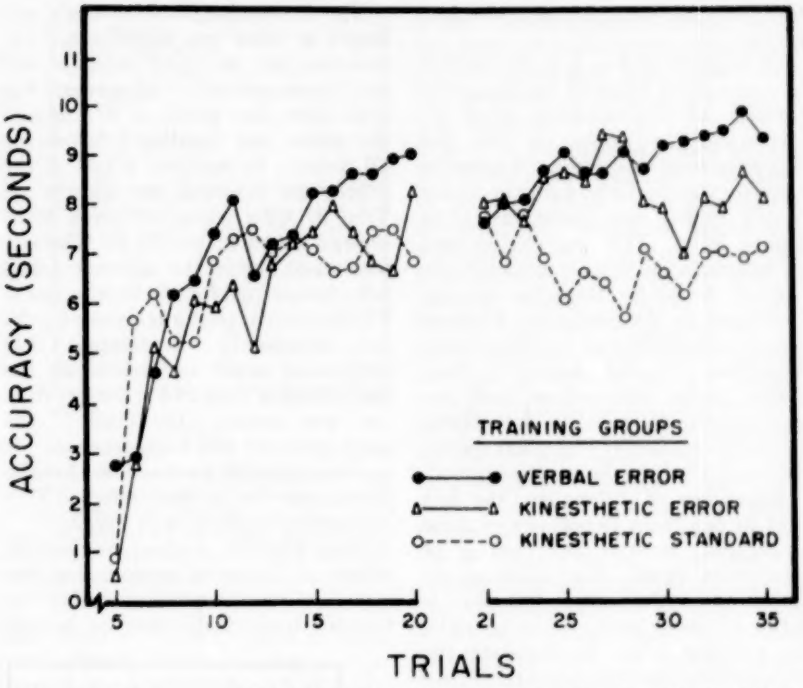

FIG. 1. Rate accuracy achieved with three kinds of knowledge of results. The break in the curves represents the rest pause. Trial 5 is the fifth prelearning trial. The maximum possible score is 15 sec.

the first three trials. On the fifth prelearning trial the average constant error for all *Ss* combined was plus 35 rpm.

Learning period.—When information about performance was provided, rate accuracy increased rapidly for all groups. These results are shown in Fig. 1 which pictures both halves of the learning period. It can be seen that the group which worked with the kinesthetic standard achieved a greater improvement in accuracy than any other group on the first learning trial (Trial 6). In fact further practice had relatively little permanent effect on the accuracy achieved by this group. In contrast, the other groups required considerably more practice in order to achieve effective utilization of the available rate cues.

During the first half of the learning period the over-all difference among groups was not significant. In the second half of this period the over-all difference was significant. These results indicate that, after sufficient practice, direct error information is more effective as an aid in learning to produce a rate of movement than is information about the standard rate itself. It should be remembered, however, that the group which worked with kinesthetic error information also received a verbal indication of the direction of error. No such indica-

tion was provided for the standard group.

Of interest in Fig. 1 is the decline in scores after Trial 28 for the group trained with kinesthetic error information. In order to test the significance of this decline, a separate analysis was performed on the scores of the verbal and kinesthetic error groups only. This analysis showed no significant difference between the over-all means for these two groups. Furthermore, the interaction between groups and trials was not significant. According to this analysis, then, verbal error information was not significantly superior to kinesthetic error information at any point during the second half of the practice period.

Regardless of differences, the fact that *Ss* can learn to use any of these ambiguous, intermittent cues is of interest in itself. The question remains as to the actual use that is made of these cues. It is possible, for example, that *Ss* translate the kinesthetic error cues into verbal cues according to the perceived magnitude of the errors.

Criterion period.—During the criterion trials *Ss* were deprived of externally administered knowledge of results. An analysis of the criterion scores indicated that neither previous training nor length of delay produced a significant over-all effect in the criterion trials. Furthermore, training method did not interact significantly with any variable or combination of variables. This resulted in spite of the fact that a significantly higher level of accuracy had been achieved by the verbal and kinesthetic error groups during the second half of the learning period. Such a condition reflects, in part, the extreme between-subject variability that appeared with this rather ambiguous criterion task.

The interaction between trials and length of delay was significant. For this analysis the error estimate was not homogeneous. However, the error term was based on 855 *df* and the effect was significant below the .01 point. In addition a test of the differences between the groups on Trial 37 (the second criterion trial) showed a significant effect. The test was made before the differences had been determined for the various trials. This particular trial was chosen for the test because it was thought that differences might be greater on the first criterion trial and a conservative test was desired. On Trial 37 the error estimate was homogeneous. It seems reasonable to conclude that the interaction is a real effect. This interaction is pictured in Fig. 2.

From Fig. 2 it is apparent that the effects of delay in reproducing the rate were most important during the first few trials of the criterion period.

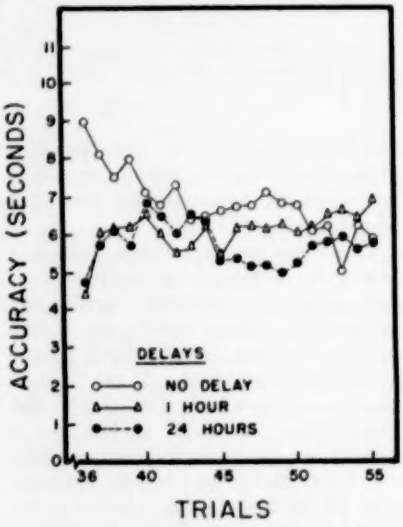

FIG. 2. Retention of rate accuracy without externally supplied knowledge of results during the criterion trials.

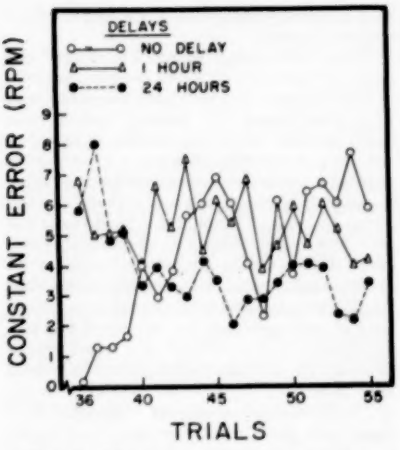

FIG. 3. Mean constant errors in rate produced without externally supplied knowledge of results. Error was calculated as the mean of the algebraic sum of all differences between the standard and the actual rate. The signs of all errors in the figure are positive.

The effects of the two lengths of delay were almost identical. Apparently there is some critical length of delay beyond which an initial reduction and later improvement in performance will appear. The critical length of delay seems to be less than 1 hr. in length.

Both of the delayed groups showed an initial rise in accuracy. For these groups the difference between the first and second criterion trials was significant. The group that experienced no delay showed a steady decline in accuracy during the early trials. These changes in accuracy appear to be associated with changes in constant errors of rate during the early criterion trials. Figure 3 shows mean constant error as a function of trials for the criterion period. Despite the variability in these points certain consistencies are apparent. The signs of the constant errors are all positive, just as they were in the prelearning trials. The magnitudes of the errors, however, are much smaller than in the prelearning trials. The *Ss* learned to reduce their rate considerably, but not quite enough.

Both of the delayed groups started at a fast rate and tended to slow down in the early trials. In contrast, the group that experienced no delay began with almost zero constant error and then increased the average rate of turning during these same trials. These consistent tendencies to increase or decrease the average rate of turning are reflected by the changes in accuracy observed in the first few trials of Fig. 2. Accuracy decreased for the *Ss* who experienced no delay because their rate showed increasing deviations from standard as trials proceeded. Accuracy increased for the delayed *Ss* because their rate was approaching the standard during the early criterion trials. Beyond Trial 40 there is little systematic relation between accuracy and constant error.

Relearning period.—When information was again supplied, all groups showed immediate improvement in performance. By the end of the third trial both the verbal error and the kinesthetic error groups had reached a level of accuracy just slightly below the highest level reached in the second half of the learning period. During the relearning period the performance of the kinesthetic error group showed no tendency to fall below the performance of the verbal error group. In contrast the performance of the kinesthetic standard group showed the same variability exhibited in the second half of the learning trials.

DISCUSSION

Retention in motor learning has been tested for a wide variety of lengths of delay between original learning and the

testing of retention. Typically a kind of warm-up effect is noted early in the testing period. It appears that the delayed groups in this experiment have shown a kind of warm-up effect. However, there is an important difference between the conditions under which retention was tested in this experiment and the conditions established in previous experiments. Bell (2), for example, was concerned with the retention of pursuit rotor skill. In each of his studies the cues available in the test of retention were the same as those available in the learning task. In the present experiment retention was tested in the absence of the cues that had been supplied in the learning period. The cues remaining were those inherent in the task itself. The warm-up effect observed in Fig. 2 apparently resulted from the use of self-supplied kinesthetic cues that *S*s had learned to appreciate.

It is clear that *S*s can learn to use externally administered kinesthetic and verbal cues in learning a rate of rotary movement. Furthermore, when they are deprived of external cues *S*s can still maintain the required rate with a fair degree of accuracy by utilizing kinesthetic cues available in the task itself. Under these conditions the final level of deterioration in performance associated with continued attempts to produce the rate after no delay is as great as the final decrement associated with a delay of as long as 24 hr.

These results were obtained with a task involving rotary movement. There is some evidence that similar results might appear with a task requiring linear movement. In a recent experiment Gottsdanker (3) found that, for short periods of time, *S*s could also maintain a constant linear rate with a fair degree of accuracy when restricted to kinesthetic cues. Apparently the rhythmical nature of rotary movement is not critical for the present results.

Summary

This experiment was designed to test the ability of *S*s to utilize kinesthetic and verbal cues in learning and retaining a constant rate of rotary movement.

The *S*s were trained to turn a handwheel at a constant linear rate of 160 ft. per min. (angular rate of 100 rpm) with information supplied after each trial during the learning period. One group received verbal information in regard to amount and direction of error; a second group received kinesthetic error information concerning the amount of rate error and a verbal indication of the direction of error; and a third group was given kinesthetic information about the standard rate. After training had been completed *S*s attempted to produce the same rate without the aid of the externally administered cues. One-third of each training group attempted this task immediately, a second third was tested after a delay of 1 hr, and the remaining third was tested after 24 hr.

The results indicated that either verbal or kinesthetic error information was superior to kinesthetic information about the standard rate when the rate was being learned. In addition, *S*s were able to maintain the rate with a fair degree of accuracy after they were deprived of the cues used in learning. The final level of accuracy achieved under these conditions was unaffected by the amount of delay experienced before the retention of the rate was tested.

References

1. Battig, W. F. The effect of kinesthetic, verbal, and visual cues on the acquisition of a lever-positioning skill. *J. exp. Psychol.*, 1954, **47**, 371–380.
2. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**, No. 1 (Whole No. 243).
3. Gottsdanker, R. M. Prediction-motion with and without vision. *Amer. J. Psychol.*, 1952, **65**, 533–543.
4. Howland, D., & Noble, M. E. The effect of physical constants of a control on tracking performance. *J. exp. Psychol.*, 1953, **46**, 353–360.
5. Lincoln, R. S. Learning a rate of movement. *J. exp. Psychol.*, 1954, **47**, 465–470.
6. Lincoln, R. S. Rate accuracy in handwheel cranking. *J. appl. Psychol.*, 1954, **38**, 195–201.
7. Weiss, B. The role of proprioceptive feedback in positioning responses. *J. exp. Psychol.*, 1954, **47**, 215–224.

(Received March 21, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

SPONTANEOUS RECOVERY AND STATISTICAL LEARNING THEORY [1]

LLOYD E. HOMME [2]

Indiana University

To be reported here are two experiments designed to uncover relationships between spontaneous recovery and (*a*) number of reinforcements, and (*b*) spacing of reinforcements. Further, the experiments are designed so that the results may be compared with predictions from statistical learning theory (**1**, **2**).

In Estes' original model (**1**) a stimulating situation is represented as a single set of elements which remains fixed during the experimental period. While this single-set notion seems adequate for dealing with many events taking place in one session, it has been found necessary, when the problems of spontaneous recovery and related phenoma (**2**) are considered, to postulate that the set of elements available during any one experimental period is a random sample from a larger total set of elements available to the organism, and further, that the elements in the experimental set are replaced, over any period of time, by random sampling from the larger set. The assumption is made throughout that the probability of occurrence of a response, R, is equal to the proportion of elements conditioned to R in the experimental subset.

To show how the model may be related to experimental events, we shall consider an initial Skinner-box learning session for a naive animal. At the beginning of the session there are, let us assume, no elements conditioned to bar-pressing. This is illustrated in Fig. 1A, by letting the larger circle, S^*, represent the total available elements, and S the portion of the total set available on any experimental period. At the end of the conditioning period all of the elements, let us say, in the experimental subset are conditioned, but, neglecting interchanges during the experimental period, none of the elements outside this subset are conditioned ones. This state of affairs is represented in Fig. 1B.

Now, if we allow 24 hr. to elapse before replacing S in the apparatus for extinction, it is apparent that the sample of stimulus elements present during the extinction period will not include all the elements that were conditioned the previous day; some conditioned elements may be thought of as having "drifted out" of the sample, while some unconditioned elements have "drifted in," as in Fig. 1C.

Now if we carry out extinction until all the elements in S are extinguished

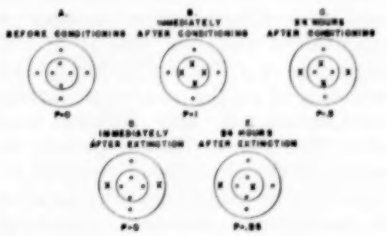

FIG. 1. Schematic representation of stimulus element situation in relation to some experimental operations. X's represent elements conditioned to R; dots represent elements not conditioned to R (conditioned to $\bar{R}$). The larger circle represents the total available population of elements; the smaller circle represents the portion of elements available on asingle experimental session.

[1] This report is a revised portion of a dissertation submitted to Indiana University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer is deeply indebted to Professor W. K. Estes for his generous advice and assistance.

[2] Now at the University of Pittsburgh.

(Fig. 1D) and then allow 24 hr. to elapse before replacing S in the apparatus, some of the extinguished elements will have drifted out to be replaced by conditioned ones (Fig. 1E). The probability of response now will be greater than at the end of the previous extinction period; spontaneous recovery will have occurred.

From the foregoing considerations it may be seen that several predictions can be made concerning the effect of *number of reinforcements and spacing of reinforcements* on spontaneous recovery from extinction. Below are statements of the pertinent predictions regarding these variables deducible from Fig. 1 on a verbal basis. Mathematical expression of these predictions will be given in a later section.

In statistical learning theory it is specified that the number of elements conditioned in a single experimental period is an increasing exponential function of the number of reinforcements, so that, in terms of Fig. 1, as the number of reinforcements is increased there would result a greater number of conditioned elements in S. Now, with the simplifying assumption that all elements have an equal and independent probability of being drawn on any one experimental period, it is clear that the number of conditioned elements present at the beginning of a first extinction period will depend on the number originally conditioned. It also follows that after the first extinction period, in which all the conditioned elements in S become unconditioned, the number of conditioned elements remaining outside S will also depend on the degree of original conditioning. If this is true, the likelihood that conditioned elements will drift back into S before the test for spontaneous recovery will also depend on the number of reinforcements during conditioning. From these considerations, then, the following predictions emerge: (*a*) initial probability in extinction should be an increasing (exponential) function of the number of reinforcements, (*b*) the level of recovery 24 hr. after a first extinction period should be an increasing function of the number of reinforcements, and (*c*) the total number of responses over the extinction series should be an increasing function of the number of reinforcements.

By the relationships of the model presented above, a given number of reinforcements spaced over several periods should result in more responses during extinction than the same number of reinforcements in one period. This prediction is a consequence of stimulus fluctuation between acquisition periods. With massed acquisition only S elements are available for conditioning; with spaced acquisition more of S^* is available since several samples of size S are drawn. Therefore, with spaced acquisition more *different* elements will be present at some time during the course of the reinforced trials and on statistical grounds we should expect more elements to become conditioned than in the case of masses acquisition. The predictions are, then, that (*a*) initial probability in extinction should be greater for the spaced groups, (*b*) there should be a higher level of recovery 24 hr. after extinction for the spaced groups, (*c*) the total number of responses over the extinction series should be greater for the spaced groups, and (*d*) there should be a greater difference between a group with five spaced acquisition periods and its massed acquisition control than between a group with two spaced acquisition periods and its control group.

These predictions are tested in the conditions of the investigations that are to be described and reported here.

Predictions from other theories of recovery.—Although these experiments were specifically designed to test a set of predictions from a statistical theory of learning, it is likely that other theories of extinction and recovery could mediate many of the same predictions. An extended review of extinction "theories" may be found in Razran (**9**) and will not be attempted here. Estes' theory of extinction, from which we will derive quantitative expressions for treatment of the present data in a later section, might be thought of as a more formal and testable statement of notions characteristic of Guthrie (**4**) and Wendt (**11**), in that the animal is considered to be learning a competing response in the presence of a situation which previously set the occasion for a conditioned response. Hull's theory, on the other hand, might be thought of as combining certain features of Pavlovian inhibition theory and Guthrian competing response theory (cf. Spence [**10**]). Because of the relatively incomplete specification of the ${}_{S}I_{R}$ construct, the design of the experiments and treatment of the data reported here have been organized in terms of the more quantitative relationships derivable from the statistical theory.

Method

Design.—Six groups of *S*s were used. Groups 1–4 received 15, 50, 100, or 250 reinforcements, respectively, during a single experimental period. These groups will be compared to evaluate the effects of number of reinforcements on extinction and spontaneous recovery. Group 5 received 100 reinforcements spaced 50 per day for two days; Group 6 received 250 reinforcements spaced 50 per day for five days. Groups 5 and 6 will be compared in conjunction with Groups 3 and 4 as in a factorial design in order to evaluate the effects of spacing, number of reinforcements, and the interaction between these two variables on extinction and spontaneous recovery.

Subjects.—The *S*s were 144 albino rats 70 to 90 days old obtained in six shipments of 24 each from the Windsor Biology Gardens, Bloomington, Indiana. One shipment consisted of males, the other five of females. The *S*s in each shipment were randomly assigned to the six experimental groups with the restriction that four go into each group. Each experimental group, then, contained a total of 4 males and 20 females.

Apparatus.—The apparatus consisted of eight identically constructed Skinner-type boxes. The rat was contained in a chamber 10½ × 12 × 10 in. having a floor of hardware cloth, side and rear panels of fiberboard covered with wire-mesh screen, and a front panel of aluminum. From the left side of the front panel a 2-in. wide bar (with the leading edge rounded to reduce the frequency of bar-biting) protruded ½ in. For this experiment the bar and rotary microswitch were adjusted so that a 4-gm. weight on the leading edge would close a circuit activating the water-dipper mechanism when the bar was depressed ⅜ in. from its original position 3½ in. from the floor. During extinction the dipper mechanism was shunted out so that only the recorder was in the circuit. Each of the experimental chambers was enclosed in a 28½ × 13¾ × 14½ in. partially soundproofed box. A 10-w. bulb mounted between the rear wall of the experimental chamber and the outer box furnished sufficient illumination so that *S* could be observed through a Plexiglas aperture cut in the top of the soundproofed outer box. Responses were recorded in an adjoining room by means of an Esterline-Angus Operations Recorder.

Procedure.—The *S*s in each shipment were assigned to four squads, each containing one member of each experimental group. All *S*s in a squad were run simultaneously in six of the eight Skinner boxes. (The other two boxes were used for a related experiment.) Counterbalancing was employed to control for box differences, three *S*s in each group being assigned to each of the eight Skinner boxes.

All *S*s were tamed by handling for at least three days prior to the experiment and were placed on a 22-hr. water-deprivation schedule for at least six days prior to experimentation. Food was available in the home cages at all times.

After taming and watering-schedule training, *S*s were conditioned to run to the dipper and drink at the sound of the "dipper click," a characteristic sound produced when the dipper solenoid retracted the dipper into the water reservoir and released it. This dipper training, carried on for 1 hr. per day for three days, was accomplished by placing *S*s in Skinner boxes in which the bars had been covered with a metal shield. The boxes were equipped with a timer which operated the dippers and a counter every 5 min. In each day's 1-hr. session, then, the dipper click followed by water presentation occurred 12 times.

Intermittent observations by *E* through the top of the box showed that all *S*s learned to approach and drink from the magazine at the sound of the click.

Conditioning.—After *S*s had received three days' dipper training, the metal shields were removed from the bars and the bar circuit was switched in so that whenever bars were depressed, the dippers and counters would operate. When the predetermined number of bar presses had been made, *S*s were removed from the apparatus and watered for 1 hr.

Extinction.—On the day following the final acquisition period, the dipper mechanism was shunted out of the circuit so that only the recorder was activated when bar-pressing occurred. After the apparatus was tested and the dippers were dried, *S*s were placed in the boxes and left there for 1 hr. At the end of this time *S*s were watered for 1 hr. after being returned to their home cages. This procedure was repeated daily for five extinction periods.

Results

The mean total extinction responses during successive 60-min. periods for Groups 1–4, where number of reinforcements during a single session was varied, are shown in Table 1. It was predicted that initial probability would differ among groups receiving differing numbers of reinforcements. Since it is specified in statistical learning theory that rate of responding is directly related to probability, the cumulated totals for the first 3 min. (by the third minute the totals have lined up in the order which is maintained over all totals) have been compared. The means (14.62, 22.67, 23.71, and 24.96 for Groups 1–4, respectively) were found by analysis of variance to differ significantly ($F = 6.43$, df 3 and 80, $P < .01$).[3] The prediction with respect to initial probability will be discussed further in connection with the p_0 parameter in the theoretical equations.

Mean response totals for the first day of extinction are in agreement with other studies of the number-of-

[3] In this instance and those to follow, wherever a significant P is reported, Bartlett's test for homogeneity of variance has yielded a nonsignificant value.

TABLE 1

Obtained and Predicted Mean Total Responses During 60-Min. Extinction Sessions on Successive Days

Group	Reinforcements	Measure	Extinction Day 1		2		3		4		5	
			Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	15	Obtained	88.5	35.1	28.9	19.9	21.3	12.6	16.9	11.7	14.0	9.8
		Predicted	88.3		34.5		23.0		17.9		15.3	
2	50	Obtained	94.7	49.6	20.1	14.8	23.9	17.0	17.9	15.3	15.1	12.8
		Predicted	94.7		35.0		23.2		17.9		15.4	
3	100	Obtained	110.8	56.8	37.5	23.0	22.0	8.9	19.1	14.6	19.4	15.1
		Predicted	110.5		35.9		23.5		18.2		15.4	
4	250	Obtained	119.1	39.7	42.1	20.1	24.6	12.3	25.3	24.7	18.1	14.4
		Predicted	119.1		36.3		23.6		18.2		15.5	
5	50 × 2	Obtained	102.2	36.9	40.9	23.9	21.0	8.4	20.0	11.9	13.8	9.4
		Predicted	102.4		35.6		23.4		18.1		15.4	
6	50 × 5	Obtained	149.9	55.5	57.1	23.8	35.5	21.8	26.2	16.4	19.3	13.9
		Predicted	150.5		36.9		23.9		18.3		15.5	

reinforcements variable (**8, 12, 13, 14**), although the F obtained in analysis of variance is not significant ($F = 2.26$; F required at .05 level is 2.72 for 3 and 80 *df*). Mean totals for the five extinction sessions (169.6, 181.7, 208.8, and 229.7 for Groups 1–4, respectively) fall in the predicted order. Analysis of variance yields an F value of 2.36; 2.72 is required at the .05 level with *df* 3 and 80.

The mean daily extinction totals for Groups 3–6, where number and spacing of reinforcements was varied, are also shown in Table 1.

Initial probabilities in terms of rate of responding were evaluated as in a factorial experiment with two variables, number of reinforcements and spacing. Neither variable yielded a significant F but the interaction term was significant at well beyond the .05 level ($F = 6.28$, 1 and 80 *df*).

It will be noted in Table 1 that, for the extended spacing group (50×5 reinforcements), there are, as predicted, more responses than for its control (250 massed reinforcements). In the case of spacing over two periods (50×2) versus one (100 reinforcements), this difference is not observed. Similar relationships are to be seen throughout the five extinction sessions. The totals over all periods are 208.8 and 197.9 for the 100 vs. 50×2 reinforcement groups and 229.7 and 288.0 for the 250 vs. 50×5 reinforcement groups. Evaluating the differences between these as in a factorial experiment, analysis of variance yields a significant F for the reinforcement variable ($F = 10.03$; $P < .01$, 1 and 80 *df*). Neither the F for spacing nor the F for interaction attains significance at the .05 level.

When the mean totals for Day 1 are analyzed in the same way, the number of reinforcements variable is significant at the .01 level ($F = 8.15$, 1 and 80 *df*), and the interaction term is significant at the .05 level ($F = 4.01$), but the spacing variable is not significant ($F = 1.29$). The significant interaction term is of particular interest because it supports the prediction of a greater difference between the five-period spacing group and its control than between the two-period spacing group and its control group. The t statistics computed between the mean total on Day 1 for Group 6 and its control, Group 4, is significant at the .05 level ($t = 2.21$, 46 *df*). The difference between Groups 3 and 5 is not significant.

Relationships Between Data and Theory

In the introductory verbal statement of the model an important simplifying assumption was made that all elements have an equal and independent probability of being drawn on each session. Quantitative predictions utilizing this assumption will be presented and compared with the data.

Let us suppose there are, altogether, C_o elements conditioned in S_* (cf. Fig. 1) at the end of conditioning. Now, on the assumption that all elements fluctuate and have equal probabilities of being drawn, there will be available, at the beginning of the first extinction period, only the fraction $\frac{S}{S_*}$ of those originally conditioned. Letting $N_{e,1}$ be the number of conditioned elements available at the beginning of the first extinction period, then

$$N_{e,1} = \frac{S}{S_*} \cdot C_o \qquad (1)$$

Now, if extinction has been carried out until probability of responding approaches zero in this session, C_o will have been reduced by this amount, $\frac{S}{S_*} C_o$, so that with a sufficient recovery period the number of elements available at the beginning of the second extinction period,

$N_{e,2}$, is

$$N_{e,2} = \frac{S}{S_*}\left(C_o - \frac{S}{S_*}C_o\right) = \frac{S}{S_*}C_o\left(\frac{S^* - S}{S_*}\right). \quad (2)$$

Letting $(1 - j) = \frac{S_* - S}{S_*}$, we may generalize Equation 2. The number of conditioned elements available at the beginning of the m^{th} session is given by

$$N_{e,m} = \frac{S}{S_*}C_o(1 - j)^{m-1}, \quad (3)$$

where m is the ordinal number of the extinction session.

Relating number of elements, $N_{e,m}$ to probability we have, letting $p_{o,m}$ be probability at the beginning of the m^{th} session,

$$p_{o,m} = \frac{N_{e,m}}{S} = \frac{C_o}{S_*}(1 - j)^{m-1}. \quad (4)$$

Now in statistical learning theory (1) an equation for the cumulated number of bar-pressing responses is given. On the assumption that extinction is the learning of competing behaviors, Equation 5 follows directly (See [5] for a fuller derivation).

$$R = \frac{1}{\theta}ln(1 - p_o) \quad (5)$$

where θ equals the proportion of elements effective at a given moment, p_o is the initial probability in extinction, and R is total number of responses during a period. Now, by substituting Equation 4 into Equation 5, we may write for the number of responses on the m^{th} day, R_m, the following:

$$R_m = \frac{1}{\theta}ln[1 - (p_o)(1 - j)^{m-1}]. \quad (6)$$

The values obtained with Equation 6 are shown in Table 1 opposite the designation, *predicted*. The numerical form of Equation 6 from which these values were obtained is

$$R_m - 12 = 28.7\,ln[1 - (p_o)\,.585^{m-1}].$$

The p_o values used were .930, .944, .968, and .976, for the 15, 50, 100, and 250 reinforcement groups, respectively; for the 50×2 and 50×5 groups the p_o values used were .957 and .991. The $\frac{1}{\theta}$ used here is the value found in fitting acquisition data from comparable animals in a different experiment using the same apparatus (4). The $(1 - j)$ is from curve fitting by approximation, the p_o values are adjusted so that the predicted and obtained first day's totals coincide, and an operant rate of 12 responses per hour has been subtracted each day to approximate the theoretical requirement that rate of responding go to zero each session. (Homme found in an independent experiment [4] with comparable animals in the same apparatus, that, for 12 days' extinction the mean operant rate was 11.67 responses per hour for the last three days.) It should be emphasized that j and θ are constant for all predictions; only p_o is permitted to vary.

It seems that the discrepancies between data and predicted curves are not of a nature to require rejection of the hypothesis that the predicted curves are of the required form and that the obtained curves have the same slope. It will not be overlooked, however, that there are systematic discrepancies between these curves and the data. For the low reinforcement groups too many responses are predicted, and for the high reinforcement groups, too few. Secondly, to ensure that the first theoretical and obtained points coincide, p_o values required are higher than the unmodified assumption of random fluctuation of stimulus elements permits; even if all of the elements in an experimental sample (S of Fig. 1) were conditioned during massed acquisiton, p_o at the beginning of extinction 24 hr. later could not exceed j, which in this case is .415.

It is worth noting that the deviations mentioned above are strikingly similar to the sort to be expected if there were a "common core" of stimulus elements having a very high probability of being drawn, which are present on *all* experimental sessions. It seems reasonable to

suppose that there are certain components of the stimulating situation which *are* almost inevitably present on every trial, e.g., some of the proprioceptive cues attendant on bar-pressing, and some which are not, e.g., the visual stimulation attendant on bar-pressing. (Although the bar must go down in a particular way on each response, this is not primarily contingent on the rat's visual orientation.) The assumption that the probability of being drawn on a given trial varies among stimulus elements has already been given explicit theoretical treatment in Estes' and Burke's discussion of the "θ distribution" (3). The common-core notion may be viewed as an instance or special case of the θ distribution hypothesis, making the simplifying assumption that the distribution is made up of two components, the very high probability elements constituting the common core, and the fluctuating elements whose probabilities are within a range such that serious error is not introduced by dealing with their mean. The consequences of the common-core assumption are that initial probability as well as total responses on the first extinction period ought to be greater than can be accounted for by the unmodified random fluctuation rule. By the same reasoning, fewer responses on succeeding sessions are to be expected to the extent that already extinguished elements of the common core continue to occupy the sample space but do not, of course, contribute to response strength.

It can be shown quantitatively that the common-core assumption will increase the correspondence between obtained and predicted total daily extinction responses and that this effect is particularly noticeable on the Day 2 values. Also the required initial probabilities will assume more theoretically reasonable values. In view of the variability exhibited in the obtained data, however, it is not clearly certain that this theoretical modification is required at the present time.

Summary

Investigations are reported of the theoretical and empirical relationships between spontaneous recovery from extinction and (*a*) number of reinforcements and (*b*) spacing of acquisition. The experiments utilized albino rats in a Skinner-type conditioning situation.

To study the effect of number of reinforcements, four groups received 15, 50, 100, or 250 reinforcements on the same day, followed by five daily 60-min. extinction sessions. It was concluded that the data offer considerable support for predictions from Estes' statistical theory of learning concerning: (*a*) initial probability of responding during extinction, (*b*) ordering of day-to-day curves of total responses in extinction, and (*c*) the convergence and slope of the curves.

To study the effect of massed versus spaced acquisition, two massed acquisition groups received 100 or 250 reinforcements on the same day and two spaced groups received 50 reinforcements on each of the two days, or 50 reinforcements on each of five days. All groups received five daily 60-min. extinction sessions. Theoretical predictions concerning higher initial probabilities and higher totals of responding during extinction following spaced acquisition were supported only by the differences between the five-day spaced acquisition group and its control. Predictions concerning the differential effect of five-day versus two-day spacing and the form and shape of the day-to-day curves of total responses in extinction are supported.

The statistical learning theory which the investigations were designed to test assumes that all elements in the stimulus population fluctuate randomly in and out of the effective stimulating situation. It was mentioned that a better quantitative account of the data could be obtained if this assumption were modified to allow for a common core of stimulus elements which remain fixed while the remainder of the population is subject to random fluctuations.

References

1. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
2. Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.
3. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
4. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
5. Homme, A. S. Rate of reconditioning as a function of amount of extinction. Unpublished doctor's dissertation, Indiana Univ., 1953.
6. Homme, L. E. Spontaneous recovery from extinction in relation to number of rein-

forcements, spacing of reinforcements, and duration of initial extinction period. Unpublished doctor's dissertation, Indiana Univ., 1953.

7. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
8. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
9. Razran, G. H. S. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264–297.
10. Spence, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 690–730.
11. Wendt, G. R. An interpretation of inhibition of conditioned reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258–281.
12. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–521.
13. Wyckoff, L. B. Resistance to extinction of the lever-pressing response in the white rat as a function of number of reinforcements. Unpublished master's thesis, Indiana Univ., 1950.
14. Yamaguchi, H. G. Superthreshold reaction potential ($_sE_R$) as a function of experimental extinction (n). *J. exp. Psychol.*, 1951, **41**, 391–400.

(Received March 31, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

"BACKWARD" LEARNING IN PAIRED ASSOCIATES

BENNET B. MURDOCK, JR.

University of Vermont

When *S* learns A—B, is he at the same time learning B—A? This possibility has been suggested in explaining the results of two recent experiments, one on mediated generalization (**4**) and one on negative transfer (**5**). Also, "backward" learning has received considerable attention in the remote backward associations of serial learning (**3**). One can argue that an A—B—C—D—E serial list is composed of successive pairs (A—B, B—C, etc.) and that the demonstrated presence of backward associations proves that the backward learning of B—A, C—B, etc. must have occurred. The study of backward learning in serial lists, however, is complicated by the facts that (*a*) each syllable functions both as a stimulus and as a response, and (*b*) adjacent "pairs" are linked together by virtue of an identical syllable.

It was felt that the problem of backward learning warranted further investigation, especially in view of its possible importance in understanding the learning process. Therefore the present study was undertaken to determine if backward learning does take place and, if so, to what extent it occurs.

Procedure

To study backward learning the present experiment used a transfer of training design which was developed to control for practice effects. The design is shown in Table 1 and was as follows: *S*s learned 10 A—B pairs on List I, and then learned List II which contained 15 pairs. The *S*s in Group I learned, in List II, five pairs (A—B) that were merely repeated from List I, five pairs (B—A) that were reversed, and five control pairs (A—B_C) that were completely new to them. If *S*s had learned B—A directly on List I, the B—A pairs on List II should show positive transfer (i.e., be easier to learn than A—B_C). The A—B pairs were included in List II to determine maximum transfer under these conditions.

Positive transfer could also result if *S*s had merely learned the individual B and A members (rather than the connections between them) on List I. If this were the cause of the positive transfer, there should also be facilitation if the B and A members were scrambled or rearranged on List II. However, the rearrangement would produce interference if the specific B—A connections had been learned on List I (**5**). Therefore, to determine whether connections or individual members had been learned on List I, *S*s in Group II learned a 15-pair List II that contained five A—B pairs rearranged (A—B_R), five B—A pairs rearranged (B—A_R), and five control pairs (A—B_C). Just as A—B would provide a standard to evaluate the degree of positive transfer for B—A, so A—B_R would provide a standard to evaluate the degree of negative transfer for B—A_R.

The *S*s were 24 undergraduate psychology students, male and female. The method used anticipation, pronounciation, noncorrection, and a varied order of presentation; no intertrial interval; and a 3-sec. presentation rate (1.5 sec. for S, 1.5 sec. for S—R).

All *S*s first learned a practice list 10 pairs long. It was composed of 20 Glaze (**2**) 93% association value nonsense syllables paired at random. The *S*s learned the list to a criterion of three correct anticipations in one trial. List I was composed exclusively of Glaze 100% syllables and was learned by all *S*s to a criterion of one perfect trial. List II was also composed only of 100% syllables, and was learned: (*a*) until each of the 15 pairs had been correctly anticipated at least once, although not necessarily on the same trial, or (*b*) for 30 trials, whichever came first.

Prior to learning List II, *S*s were merely told that the list would contain both old and new nonsense syllables. If they asked if the pairs would be the same (or different), they were told to wait and see. Approximately 3 min. elapsed between the end of List I and the start of List II.

List II required 30 100% syllables; these were selected from the Glaze list so as to minimize duplication of letters and letter combinations. The 30 were randomly paired to form the Group

TABLE 1
DESIGN OF EXPERIMENT

List I Groups I & II (N = 24)	List II					
	Group I (N = 12)			Group II (N = 12)		
A—B	A—B	B—A	A—Bc	A—B_R	B—A_R	A—Bc
A_1—B_1	A_1—B_1	B_6—A_6	A_{11}—B_{11}	A_1—B_4	B_6—A_6	A_{11}—B_{11}
A_2—B_2	A_2—B_4	B_7—A_7	A_{14}—B_{12}	A_2—B_2	B_{10}—A_7	A_{12}—B_{12}
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
A_{10}—B_{10}	A_5—B_5	B_{10}—A_{10}	A_{15}—B_{15}	A_5—B_1	B_5—A_{10}	A_{15}—B_{15}

I list, but pairs with identical letters in the two syllables were ruled out. The 15 Group I pairs were rearranged as necessary for Group II. Since the basic comparisons were between groups of pairs within one list, it was necessary to control for possible differences in ease of learning among the individual S—R pairs. Both Group I and Group II were broken down into three subgroups. With three experimental conditions (forward, backward, and control) a 3 × 3 latin square was used so that each S—R pair appeared in a different experimental condition for each subgroup. With this counterbalancing each pair appeared equally often in the forward, backward, and control groups. There were four Ss in each subgroup, and they were randomly assigned.

RESULTS

The measure of learning used was the total number of correct anticipations per pair (abbreviated CA) for the blocks of five nonsense syllables treated alike. All computations of the significance of differences between blocks of nonsense syllables within the same lists took into account the correlation of scores, with $df = 11$. For Group I ($N = 12$) the mean difference between A—B pairs and A—B_C pairs was +3.2 CA's, and this difference was significant ($t = 4.95$, $P < .01$). The mean difference between B—A pairs and A—B_C pairs was +2.4 CA's, also significant ($t = 2.53$, $P < .05$). The difference between A—B pairs and B—A pairs was, however, not significant ($t = 1.65$, $P > .05$). Thus, both the A—B pairs and the B—A pairs showed positive transfer, but did not differ significantly from each other.

For Group II ($N = 12$) the mean difference between A—B_R pairs and A—B_C pairs was − 3.6 CA's, and the difference between B—A_R and A—B_C pairs was − 5.1 CA's. In both cases the differences were significant (t's of 3.15 and 3.56, respectively, $P < .01$ for both), but again the difference between them was not significant ($t = 1.62$, $P > .05$). Thus, both A—B_R and B—A_R showed negative transfer; the latter actually showed more interference than the former, but this difference was not significant.

The data for Group I can give a measure of transfer for B—A. If the number of CA's for the control pairs represents zero transfer and the number of CA's for the forward pairs represents 100% transfer, the CA's for the backward pairs can be translated into a percentage of maximum possible transfer (1). With this method the B—A pairs showed a positive transfer of 73%. This figure, however, is somewhat arbitrary. It would probably have been lower, for instance, with a less stringent List I criterion, but higher with a more stringent List I criterion.

Since the B—A pairs showed positive transfer, it would appear that *Ss* had learned B—A on List I. Since the B—A_R pairs gave negative transfer, it would appear that *Ss* had learned specific backward associations rather than just the individual A and B members. As final proof of this, the difference between the facilitation of B—A and the interference of B—A_R was highly significant ($t = 4.44$, $P < .01$).

Discussion

Under the conditions of this experiment *Ss*, in learning A—B, were evidently also learning B—A. Furthermore, to judge by the amount of positive transfer shown, they seemed to be learning B—A quite well—almost as well, in fact, as they learned A—B. Actually, the data cannot disprove a hypothesis of no difference between forward and backward learning.

The finding of so much backward learning is especially surprising in view of the various conditions that would appear to be unfavorable for it. On List I everything indicated a forward sequence. The *Ss* were always shown A and asked to anticipate B, never the reverse. When A and B appeared together it was always, reading from left to right, first A then B. The *Ss* were given no warning that a reverse order would ever be used. In learning they often called out the A syllable and attempted to anticipate B; failing, they would verbalize B when it appeared. No *S* reported actively rehearsing B—A; actually, with the rapid presentation rate and absence of any intertrial interval there was little opportunity for rehearsal of any kind. On the other hand, though, use of 100% syllables may have contributed to the backward learning, especially insofar as the syllables lent themselves to mediating associations.

Clearly, *Ss*' learning in this experiment was bidirectional. But is it correct to call the B—A learning "backward"? The test for learning in List II was clearly backward with respect to the test for learning in List I, but this does not necessarily mean that the List I learning was backward. The two syllables were presented simultaneously to *Ss*; presumably they had to respond to them sequentially, but the time interval between the two responses would be very slight. Perhaps the temporal contiguity was the important factor, and which syllable came first and which second was relatively unimportant. In any case, it would seem that further work on this problem is indicated.

Summary

The experiment attempted to determine whether *Ss*, in learning A—B, also simultaneously learn B—A. Following an A—B paired-associate list, half the *Ss* showed considerable positive transfer in learning B—A directly, while the other half showed considerable negative transfer in learning "scrambled" B—A pairs. Presumably both the positive and negative transfer resulted from an initial learning of B—A. It was concluded, therefore, that backward learning does occur, although the term "backward" may not be appropriate to describe this phenomenon.

References

1. Gagné, R. M., Foster, H., & Crowley M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.
2. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
3. McGeoch, J. A., & Irion, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
4. Murdock, B. B., Jr. The effects of failure and retroactive inhibition on mediated generalization. *J. exp. Psychol.*, 1952, **44**, 156–164.
5. Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

(Received March 31, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

VERBAL LEARNING AND RETENTION AS A FUNCTION OF THE NUMBER OF COMPETING ASSOCIATIONS[1]

ROBERT S. BEECROFT

State University of Iowa[2]

Most students of verbal learning assume that one of the factors determining the probability of a correct anticipatory response to a given stimulus is the number of competing associations attached to the stimulus. Such competing associations have been given a role in theoretical treatments of the acquisition, transfer, and retention of verbal responses (**3, 8, 10, 12**).

One way of varying the number of competing associations is to utilize the development, through stimulus generalization, of generalized associations to similar stimuli within a list, and to relate the speed of learning of such a list to the number of competing associations. That verbal responses do generalize to similar stimuli, where similarity is variously defined, is indicated by a number of studies (**2, 4, 6, 11, 23**). It is presumed that when a correct association A_1-B is strengthened, that associations A_2-B, A_3-B, etc., are likewise strengthened where A_2, A_3, etc., are stimuli similar to A_1. Now if A_2, A_3, etc., are not included in the same list with A_1, such tendencies will not be revealed until *S* is specifically tested with these stimuli. If, however, A_2 is in the same list with A_1-B, and A_2 is paired with C, then presumably whenever A_2-C is strengthened, A_1-C is also strengthened through stimulus generalization. It is in this way that the competing association A_1-C is developed.

There have been two studies using this technique with paired associates that are specifically analyzable in terms of stimulus generalization. Gibson (**5**) used four sets of three similar nonsense figures as stimuli with nonsense syllables as responses in a list of 12 pairs. In this way, for any pair, presumably two competing associations would be developed. She had previously shown that a high degree of stimulus generalization would be obtained with these sets of stimuli (**4**). A control group learned the same nonsense-syllable responses to a standard set of dissimilar nonsense figures. Marked retardation was shown in learning responses to the similar stimuli.

Underwood (**18**) varied the number of nonsense-syllable stimuli with identical first and last letters. He found that trials to learn increased as a function of the number of pairs having similar stimuli.

The results of a number of other studies (**17, 18, 21, 22**) of serial and paired-associate learning cannot be interpreted in terms of the type of generalization involved in establishing the competing associations, since in these studies opportunity has been allowed for the simultaneous action of stimulus generalization, response generalization, and stimulus-response generalization. These studies, however, have consistently shown inter-

[1] This report is based on a dissertation submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to thank Dr. Harold P. Bechtoldt for directing the investigation, and the other members of his committee for their suggestions in planning the study.

[2] Now with the Human Resources Research Office, The George Washington University.

ference in learning to be a function of the number of competing associations.

Interest in retention as a function of intralist similarity has stemmed in part from Gibson's theorizing (**3**). Gibson assumed that during practice the strength of the generalized responses to similar stimuli in a list would tend to be reduced through the mechanism of differential reinforcement, and that following practice these generalized responses would tend to recover, so that recall would be poorer for lists with a high degree of intralist similarity. Subsequent research has not supported this prediction, the general conclusion being that intralist similarity does not affect recall (**5, 17, 18, 19, 20, 21**).

The purpose of the present experiment is to relate the learning and retention of paired adjectives to the number of competing associations, where the number of competing associations is manipulated by varying the number of synonymous adjective stimuli in a list. Synonyms are known to manifest stimulus generalization (**23**), and interference effects, predictable on the basis of generalization of one type or another, have been obtained with synonymous adjectives (**17, 20, 21, 22**).

Method

Lists.—The experimental lists were (*a*) no synonymous stimuli, no competing associations; (*b*) four sets of two synonymous stimuli, one competing association per pair; (*c*) two sets of four synonymous stimuli, three competing associations per pair; and (*d*) one set of eight synonymous stimuli, seven competing associations per pair.[3] These lists are shown in Table 1. The stimuli in each row (and thus in each list) are matched by their rank in thousandths in the Thorndike-Lorge word frequency count (**13**). The stimuli in the lists are composed of approximately the same number of letters and the same number of syllables. The same two-syllable adjectives were used for the responses in all lists. Within any list there is no duplication in initial letter of the words used. It should be noted that one pair, GIGANTIC-YEARLY, is common to all lists.

[3] The number of competing associations refers only to the competing associations presumed to be established by the stimulus variations employed in these lists.

TABLE 1

Pairs Used in the Experimental Lists

Stimuli				Responses All Lists
List A	List B	List C	List D	
gigantic	gigantic	gigantic	gigantic	yearly
drowsy	massive	massive	massive	jolly
remote	remote	immense	immense	fragrant
idle	distant	vast	vast	kindly
energetic	energetic	energetic	colossal	wintry
lonely	lively	lively	enormous	normal
quiet	quiet	busy	large	outer
tender	silent	active	huge	pious

Procedure.—After receiving appropriate instructions, all *Ss* were given eight trials on a practice list. The practice list was eight pairs of four-letter nouns selected from the thousand most frequent words in the Thorndike-Lorge count. Following the practice list, *S* learned his particular experimental list to a criterion of seven out of eight responses correct, by the anticipation method. The *S* was then dismissed with instructions to the effect that there would be more lists to learn on the following day. Twenty-four hours later recall measures were obtained and *S* relearned his list to the 7/8 criterion. The order of pairs used for *S* on the first relearning (recall) trial was different from the order of pairs on his criterion trial the day before, and the order of presentation was systematically varied among the four orders thereafter.

All lists were presented on a Hull-type drum, the stimulus appearing for 2 sec. followed by the joint presentation of stimulus and response for 2 sec. Four orders of each list were used. The orders were arranged so that no pairs occurred in the same adjacent sequence more than once, and so that in Lists B and C no more than two synonymous stimuli occurred in succession within any one order. The intertrial interval for all lists (including the practice list) was 4 sec.

Subjects.—The *Ss* were 40 men and 40 women drawn from two classes in introductory psychology. All *Ss* used were naive with respect to participation in verbal learning experiments and were from the middle 60% of students on the forced-choice form of the Taylor Anxiety Scale (**7**). Ten men and 10 women learned each experimental list. Six other *Ss* were run but their data were eliminated. Of these, three

failed to return for recall on the second day, and one *S* was discarded for failure to comply with instructions. The remaining two *S*s were discarded for failure to learn. One of these *S*s was attempting serial learning of responses, and the other *S* reported she was not trying to learn.

Results

Learning.—In Fig. 1, the mean number of correct anticipations on Trials 2 (first anticipation trial), 3, 5, and 7 is shown for the four experimental lists. The means have been subjected to an average adjustment based on an analysis of covariance of the mean number of correct responses on Trials 2–8 as a function of the total number of correct responses on the practice list. This corrects the set of curves for initial differences on the control variable. The learning of the lists with competing associations (Lists B, C, and D) is clearly slower during the early trials.

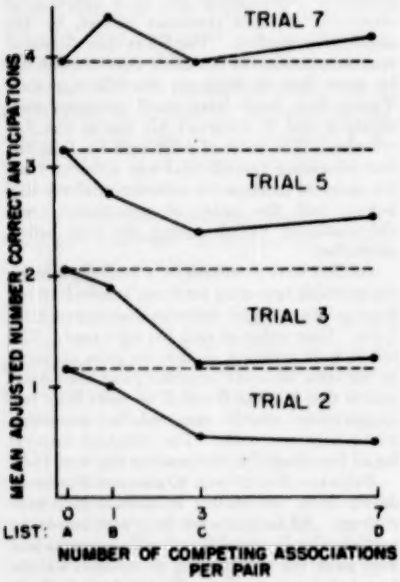

Fig. 1. Mean adjusted number of correct anticipations on Trials 2, 3, 5, and 7, as a function of the number of competing associations.

TABLE 2

Trial on Which First Correct Anticipation of the Common Pair Occurred

Measure	List			
	A	B	C	D
Mean	3.95	5.95	9.20	7.70
SD	1.94	3.38	6.39	2.58

For the pair common to all lists, GIGANTIC-YEARLY, an analysis was made of the trial on which the first correct anticipation of this pair occurred. These means for the different lists are shown in Table 2. The differences between the means are significant at the 1% level ($F = 5.87$, $df = 3$ and 72). If the Lists component in the analysis of variance is broken into two components, the one representing the contrast between List A and the other lists is significant at the 1% level ($F = 11.56$, $df = 1$ and 72) and the other representing the differences between Lists B, C, and D is significant at the 5% level ($F = 3.43$, $df = 2$ and 72).

These and other analyses indicate interference early in learning for lists in which there are competing associations. With various measures of early learning, the only contrast that falls short of significance in all analyses is between Lists C and D. All other contrasts are significant in some analysis.

The adjusted mean trials to learn to the 7/8 criterion, using the total number correct on the practice list as a control variable, and the means of the mean overt errors per trial are shown in Table 3. Although Lists C and D took longer to learn, the Lists component in the analysis of covariance is not significant ($F = 1.46$,

TABLE 3

LEARNING MEASURES

List	Adjusted Trials to 7/8 Criterion		Mean Overt Errors Per Trial	
	Mean	*SD*	Mean	*SD*
A	14.71	9.05	1.10	0.56
B	14.11	4.17	1.12	0.57
C	19.67	11.70	1.17	0.56
D	16.86	8.09	1.24	0.74

TABLE 4

RETENTION MEASURES

List	Mean Pairs Recalled		Overt Errors at Recall		Mean Trials to Relearn	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
A	4.25	1.62	0.75	0.76	4.45	3.47
B	3.90	1.04	1.05	0.79	3.55	1.32
C	3.80	1.05	1.05	0.83	3.95	1.15
D	4.00	1.69	1.15	1.35	3.20	1.31

$df = 3$ and 71).[4] While the mean number of overt errors per trial increases with the number of competing associations per pair in the list, the F obtained in the analysis of variance is less than 1.0.

Retention.—The retention results are summarized in Table 4, where the mean number of responses recalled, the mean number of trials to relearn to the 7/8 criterion and the mean number of overt errors at recall is given. The different lists were recalled equally well ($F < 1$) and there were no significant differences between lists in relearning ($F = 1.23$, $df = 3$ and 72). More overt errors were made in recall of the lists with similar stimuli, but as the total number of errors on which the means are based is small, no reliance can be placed on these data.

[4] Analysis of covariance was applied as a means of increasing the precision of the experimental results and obtaining improved estimates of the list means, wherever the data merited such an analysis. The F's obtained in the various analyses are actually less than the F's obtained with the usual analysis of variance of the data. The larger F's with the raw data are due to the relatively poor performance of *Ss* assigned to List C on both the practice and experimental lists, the covariance analysis reducing the mean square for lists.

DISCUSSION

Learning.—The results are in general agreement with previous studies of the effect of competing associations on the learning of a new association. A single competing association per pair has little effect on learning (**1, 9, 14, 16**). If there is any interference in learning, it is observed only in the first few trials. Overt errors are usually present in increased numbers with one competing association per pair.

With three or four competing associations there is usually considerable interference in the early stages of learning, this interference lasting somewhat longer than with a single competing association per pair. The frequency of overt errors remains at about the same level as with a single competing association (**19**). Maximal interference seems to be reached with three or four competing associations per pair in that a greater number of competing associations does not result in more interference.

That *Ss* showing an initial decrement due to competing associations catch up quickly with other *Ss* is a fairly general finding (**14, 15, 16**). In the present situation, it is conceivable that although the competing associations may continue to grow in strength during practice through stimulus generalization, the correct associations are growing at a much faster rate, and that once the difference in strength passes a certain minimum there is no further interfering effect of the competing associations.

The general hypothesis may be advanced that the conditions leading to decremental effects in verbal learning are those that increase the number and/or strength of competing associations relative to the strength of the correct

association. The strongest argument against such a hypothesis would be based on Underwood's recent study of response similarity (**19**). He found that the mean number of trials to a 10/10 criterion was not affected by variation in the number of pairs having similar responses, although the mean number of overt errors increased directly with this variable. Presumably, variation in the number of pairs having similar responses would lead, through response generalization, to variation in the number of competing associations per pair. It is possible that although trials to a 10/10 criterion do not vary as a function of this variable, a further examination of the data would reveal interference in the early stages of learning for the lists with similar responses. If interference were to be found in the early trials, the generalization advanced concerning the role of competing associations in verbal learning would encompass considerable data.

Retention.—The failure to find a variation in recall as a function of intralist similarity confirms the previous findings (**5, 17, 18, 19, 20, 21**). As in any theoretical scheme, several assumptions are involved in Gibson's deduction that recall should vary inversely as a function of intralist similarity (**3**). The principal assumptions are: (*a*) incorrect associations subtract from correct associations to determine the probability of a correct response; (*b*) increasing the number of similar stimuli in a list increases the number of incorrect tendencies through stimulus generalization; (*c*) these generalized response tendencies are weakened, through differential reinforcement, during learning; and (*d*) these generalized response tendencies recover strength with the passage of time since learning.

There is little reason for questioning the first two assumptions, as much data is integrated by means of these assumptions. As for the third assumption, it is true that if learning is carried to one perfect trial, the frequency of overt errors will reach zero. It is possible, however, that the error tendencies are not reduced absolutely during learning, but that as the correct response tendency gains in strength, these error tendencies become relatively inconsequential in determining the probability of a correct response. Such an interpretation is consistent with the transitory interfering effects of competing associations. As for the fourth assumption that generalized responses recover strength in time, it is apparent that if they do so they exert little or no effect on recall.

Summary

Previous studies of verbal learning have indicated that interference in learning increases with the number of competing associations. Four paired adjective lists, varying in the number of competing associations per pair, were learned by the anticipation method and recalled 24 hr. after learning. The results agree with previous findings that competing associations handicap performance early in learning and that intralist similarity does not affect recall.

References

1. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343–361.
2. Gibson, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **24**, 237–253.
3. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
4. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
5. Gibson, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.
6. Gulliksen, H. Transfer of response in human subjects. *J. exp. Psychol.*, 1932, **15**, 496–516.
7. Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447–455.
8. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univ. Press, 1940.
9. Kline, L. W. An experimental study of associative inhibition. *J. exp. Psychol.*, 1921, **4**, 270–299.

10. McGeoch, J. A. Studies in retroactive inhibition: VII. Retroactive inhibition as a function of the length and frequency of presentation of the interpolated lists. *J. exp. Psychol.*, 1936, **19,** 674–693.
11. McKinney, F. Quantitative and qualitative essential elements of transfer. *J. exp. Psychol.*, 1933, **16,** 854–864.
12. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53,** 173–203.
13. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1944.
14. Underwood, B. J. Associative inhibition in the learning of successive paired-associate lists. *J. exp. Psychol.*, 1944, **34,** 127–135.
15. Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59,** No. 3 (Whole No. 273).
16. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39,** 24–34.
17. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42,** 153–161.
18. Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intra-list similarity. *J. exp. Psychol.*, 1952, **44,** 80–87.
19. Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45,** 133–142.
20. Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjective lists as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45,** 143–149.
21. Underwood, B. J. Studies of distributed practice: X. The influence of intralist similarity on learning and retention of serial adjective lists. *J. exp. Psychol.*, 1953, **45,** 253–259.
22. Underwood, B. J., & Goad, D. Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, 1951, **42,** 125–134.
23. Yum, K. S. An experimental test of the law of assimilation. *J. exp. Psychol.*, 1931, **14,** 68–82.

(Received April 5, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

IMPORTANCE OF OBJECT RECOGNITION IN SIZE CONSTANCY [1]

ROBERT C. BOLLES AND DANIEL E. BAILEY

University of California

It has generally been taken for granted that the perception of the sizes of things is a process mediated by the traditional "cues," i.e., parameters of distance (parallax, convergence, etc.) coupled with retinal image size. It is assumed that these cues carry the information that is used in perceiving sizes of things, and that no other information is needed or used. The thesis of this paper is that, in a familiar situation, *S* uses sensory information (from the visual stimulus) to recognize and identify the "stimulus object," but that the information utilized in judging sizes does not come exclusively from the visual cues to size, but comes also from *S*'s own past experience and learning. We suggest that where size constancy is most evident, i.e., in a familiar situation, it is probably not so much the perceived size which is invariant with distance as it is the identification of the object itself.

The ability of *S* to achieve object constancy has been emphasized by Brunswik. He reports a correlation of .994 between *S*'s estimate of object size and the measured size (**1**). He found further that there was better achievement of object size than there was of the cues supposedly mediating the achievement of size. From these results Brunswik concludes that *S* had gained "command of a particular set of mediational cues and sensory instruments" by which there is a "natural focussing of the perceptual system upon the distal stimulus variables" (**1**, p. 46). We would hold that much of the information used in judging the size of a familiar object comes from *S*'s past experience with the object. We suggest that poor achievement of the visual cues to size is a reflection of the fact that *S* has relatively little need of such cues. The human adult has many habits of language and action which enable him to function more economically on the basis of the identification of the object rather than on the basis of the cues to size. The information from the visual cues in any particular situation may be redundant. In other words, he is not dependent upon immediate information *from* his environment but need only have information *about* his environment.

In the study to be reported here, *S*s made two sets of size judgments of a representative sample of familiar objects: nonvisual estimates based only on verbal descriptions of the objects, and visual judgments under free, open-field conditions. Our hypothesis is that when identity of the objects can be established verbally, *S* can dispense with sensory cues and still make good estimates of size. In support of this hypothesis this experimental situation should provide three lines of evidence. We predict:

1. Size constancy as measured by a correlation coefficient between measured size and the nonvisual judgments of size will approach unity.

2. The errors *S* makes in nonvisual judgment of size will not be, typically,

[1] This work was stimulated by remarks made by Dr. E. R. Hilgard at the Berkeley Conference for the Unity of Science (*Psychol. Rev.*, 1955, **62**, 226–228).

significantly larger than his errors in visual judgment.

3. In those cases where *S* can improve his judgment with the additional visual information, it will be because the object was not properly identified for the nonvisual judgment.

In the nonvisual judgments of size, identification is assumed to be a function of two things: first, the homogeneity of sizes of objects like the described object, and secondly, *S*'s familiarity with objects like the described object.

Method

Subjects.—The *S*s were five male psychology students.

Procedure.—Fifty-four objects in and around one of the *E*'s home were selected as stimulus objects. The logarithms of the sizes were normally distributed with a mean of 7 in. and an *SD* of 1.2 log cycles. This range is slightly smaller than that in Brunswik's sample. A test of normality yields a p of .85 that the distribution does not depart from normality. The objects included such things as pencils, light fixtures, furniture, books, articles of clothing, automobiles, lawn mowers, garden hose, print, etc. The order of presentation of the objects in both the testing and control samples was random.

The *S*s were invited to *E*'s home individually for an afternoon. The whole procedure was interspersed with leisurely conversation. Each *S* was asked to close his eyes and to guess the size of an object that *E* described to him. When *S* closed his eyes, he was told the genus and species of the object in question. For example, *E* said, "I have an ash tray. It is a ceramic ash tray with V-shaped notches in it. What is its diameter?" Reference to the size of the object was explicitly avoided. After making this judgment, *S* was told to look at the object and to make a second judgment of its size. The *S*s were also asked to make judgments of the sizes of 15 control objects while looking at them without having made preliminary nonvisual judgments.

Each *S* made a Q sort for the objects according to his own *manipulative familiarity* with them (i.e., the relative amount of handling experience he had had with objects like the verbally described object). The *E*s made a Q sort for the objects according to their *homogeneity* (i.e., the relative invariance of the sizes of the objects in that class of which a given stimulus object belonged). Thus an object described as "*Life* magazine" would be rated high, while an object described as "a yellow pencil" would be rated low in homogeneity.

TABLE 1

Correlations of Visual and Nonvisual Judgments with Measured Size

Judgment	Subjects					Mean
	1	2	3	4	5	
Visual	.995	.993	.993	.995	.993	.994
Nonvisual	.992	.981	.988	.993	.985	.988

Results

The correlations of both the nonvisual judgments and the visual judgments with the measured sizes are given for each *S* in Table 1. All of these correlations approximate unity. Interpreting these correlations as measures of size constancy, it follows that these *S*s can achieve essentially the same degree of size constancy with or without visual information. This confirms Prediction 1.

The distribution of the differences between visual error and nonvisual error, given in Fig. 1, may be taken as a distribution of improvement scores, the improvement being attributable to the additional visual

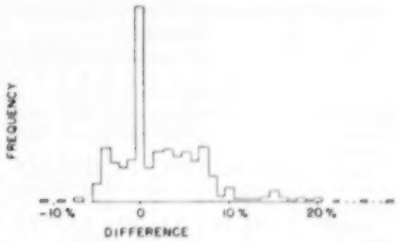

Fig. 1. Distribution of the differences between nonvisual errors and visual errors.

information furnished to *S* by presenting him with the object.

In 52% of the cases, *S*s were not able to improve their judgments with the additional visual information! In only a small proportion of the cases did the additional information enable *S*s to make significantly better judgments. This small proportion of cases contains some very large improvements; because of the skewness of the distribution, the mean is significantly greater than zero. Thus, the mean visual error is significantly smaller than the mean nonvisual error, and also the differences in correlations in Table 1 are significant. The important thing is that these significant differences are caused by relatively few judgments, which confirms Prediction 2.

Table 2 presents the mean absolute percentage error for each *S* made in visual and nonvisual conditions and in the control situation (visual judgments which were not preceded by nonvisual judgments). Percentage errors are relative to measured size and are taken without regard to sign.

It is clear from Table 2 that making nonvisual judgments prior to making visual judgments does not influence the latter. Table 2 also indicates the extent of individual differences between *S*s.

TABLE 2

MEANS OF THE ABSOLUTE PERCENTAGE OF ERRORS FOR EACH *S*

Errors	Subjects					Composite *S*
	1	2	3	4	5	
Nonvisual	16.5	23.9	16.9	14.7	26.4	19.6
Visual	12.1	15.3	10.5	8.7	11.4	11.6
Control visual	11.9	16.0	10.7	7.6	12.4	11.7
t(nonvisual -visual)	1.9	2.6	2.3	3.1	4.3	5.1
t(visual-control)	.1	.1	.1	.4	.2	

TABLE 3

CORRELATIONS OF ERRORS AND IMPROVEMENT WITH THE DIMENSIONS OF IDENTIFICATION FOR THE COMPOSITE SUBJECT

Errors	Correlation		
	Manipulative familiarity	Homogeneity	Multiple
Nonvisual	−.50	−.32	−.54
Visual	−.21	+.09	−.06
Improvement (Difference in errors)	−.49	−.39	−.56

The mean visual error, mean nonvisual error, and mean manipulative familiarity were found for each object and averaged over the five *S*s. The intercorrelations of the mean error scores and the homogeneity ratings based on the consensus of *E*s are reported in Table 3.

The nonvisual errors show a significant negative correlation with both manipulative familiarity and homogeneity, whereas the visual errors are not significantly correlated with either. Since improvement on the basis of visual information is significantly negatively correlated with both manipulative familiarity and homogeneity, we conclude that such improvement is a function of identification as it has been considered here. This confirms our third prediction.

Except for three items of print which were consistently overestimated, there was no systematic relationship between the size of object and error. These small items of print produced significant negative correlations in three *S*s.

DISCUSSION

These results show that *S* is not dependent upon visual stimulation in

making accurate estimates of the sizes of familiar objects. This fact is compatible with the findings of recent studies pointing out the role of cognitive processes in perception (**2, 3**). The correlations of the errors of estimation with the familiarity and homogeneity of the objects indicate that the tighter the sizes of a class of objects cluster around a fixed size, and the more familiar a person is with the class of objects, the better he can judge the size of a member of the class.

Our hypothesis asserts that if verbal information provides full identification, visual cues cannot elicit better size judgments. Identification was assumed to be a function of two dimensions, and an attempt was made to quantify these. However, identity depends on other dimensions also; for example, individual verbal habits, frames of reference, interest in the objects, etc. The fact that even our impoverished combined measures of identity correlate −.56 with improvement suggests that a more sensitive measure of identity would reveal that it is the primary process mediating accuracy of estimation. This assertion is not unreasonable in that size is one of the basic criteria of classification of things in our world, one of the ways in which we know things. The word *cigarette* connotes (among other things) the size of cigarettes. Thus if a thing is a foot long, it can't be a cigarette, and conversely. If a person is able to identify an object according to some classification, he has delimited its size.

It should be obvious that if a person is trained to talk about objects in his environment with certain words, he can then describe these objects with those words. If part of learning about objects is learning to talk about their sizes, then one should be able to talk about their sizes without "perceptual" mediation.

Our results indicate that the adult human has available sufficient information about his environment that, most of the time, his achievement of size constancy is independent of sensory cues to size. One might be tempted to say that in the case of familiar objects size constancy is nothing but a manifestation of stimulus error.

Summary

Two sets of judgments of size of a representative sample of familiar objects were obtained: nonvisual estimates based only on verbal descriptions of the objects, and visual judgments of the same objects made under free, open-field conditions. Correlations of estimates of size with measured size approach unity, indicating nearly perfect size constancy for both kinds of judging conditions. There is an improvement, in terms of average error, with the additional visual stimulus. However, this improvement is shown to be a function of the better identification of the objects that is possible. This questions the assertion that the perceiver is "focused" upon the distal stimulus. Most of the time, size constancy in regard to familiar objects is just as high without any mediation from the visual system as it is with the visual system available.

References

1. Brunswik, E. Distal focussing of perception: size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, **56**, No. 1 (Whole No. 254).
2. Hastorf, A. H. The influence of suggestion on the relationship between stimulus size and perceived distance. *J. Psychol.*, 1950, **29**, 195–217.
3. Hochberg, C. B., & Hochberg, J. Familiar size and the perception of depth. *J. Psychol.*, 1952, **34**, 107–114.

(Received April 7, 1955)

Journal of Experimental Psychology
Vol. 51, No. 3, 1956

EFFECT OF DELAY OF KNOWLEDGE OF RESULTS ON LEARNING A MOTOR TASK [1]

JOEL GREENSPOON AND SALLY FOREMAN

Florida State University *Pomona College*

The delay of reward has been demonstrated in a number of studies with animals (**2, 4, 5**) to slow the rate of learning. These studies have shown that as the time interval between the response and the reward is increased, the rate of learning is decreased. Saltzman (**6**) has investigated the effect of delaying reward on the rate of learning in human *Ss*. The results of his experiment are in agreement with the results obtained with animals.

Since delay of reward has been found to influence significantly the rate of human verbal learning and the rate of animal learning, it was felt that a similar effect may be demonstrated by delaying reward in a human motor learning situation. It was the purpose of the present study to investigate this problem. Specifically, this experiment was designed to study the effect of delaying knowledge of results on the speed of learning a motor task by human *Ss*.

METHOD

Apparatus.—The experiment was conducted in a small room in the psychological laboratory of Pomona College. The room contained two chairs, one on either side of a small table. A strip of glass, 2 ft. by 2 ft., was fastened to the top of the table. An L-shaped strip of metal, 18 in. long, was placed on its side across the glass and fastened to the top of the table. The metal strip was ground to a smooth edge. Lines 2¼, 3, and 3¼ in. from the left-hand stop were marked on the top of the metal strip. A roll of heavy brown paper was attached to a roller under one side of the table. The paper passed over the glass and under the metal strip to the opposite side of the table where it was attached to another roller. Thus the paper could be rolled from one roller to the other. The entire apparatus was covered with a black cloth except when in use. A mechanical pencil, a blindfold, and a stop watch were set on a ledge near the table.

Subjects.—The *Ss* were 40 undergraduate students at Pomona College who were randomly assigned to five groups of eight *Ss* each.

Procedure and design.—Each *S* was run individually. After entering the experimental room, *S* was seated at the table and given the following instructions. "What I want you to do is to draw 3-in. straight lines horizontally while blindfolded, like this [*E* demonstrated by drawing a horizontal line in the air]. When you have completed your line, leave your hand at the point where you stopped. I shall return your hand to the starting point. Also keep your arm and hand off the table while drawing your line. You are to have only the point of the pencil touching the paper. Are there any questions?" Pertinent parts of the instructions were re-read to *S* if questions were asked. If there were no questions, *S* was blindfolded, the apparatus uncovered and the pencil placed in *S*'s hand. The pencil was placed at the starting point, the left-hand stop of the metal strip, and *S* drew the line along the edge of the smoothly ground metal strip.

The information given to *S* was "long," "short," or "right." A "long" response was a line more than 3¼ in. in length. A "short" response was a line less than 2¾ in. in length. A "right" response was line between 2¾ and 3¼ in. long.

Each *S* drew 50 lines with 30 sec. between each trial.

For *Ss* in Group I, the information was given immediately on completion of the line. As soon as *S* drew his line, *E* gave the appropriate information and started the stop watch. After 30 sec. had elapsed, *E* picked up *S*'s hand, rolled up the paper and replaced *S*'s hand at the same starting point for the next trial. The same procedure was followed for all trials and for all *Ss* in all groups except for the interval between completion of the line and the pres-

[1] Portions of this paper were presented at the meetings of the Western Psychological Association by the junior author, Long Beach, Calif., May, 1954.

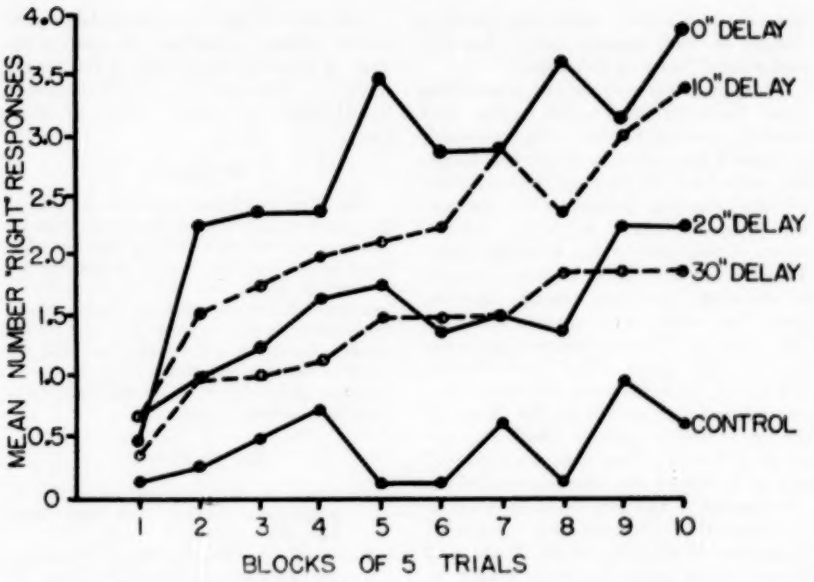

FIG. 1. Mean number of "right" responses of the control and experimental groups for successive blocks of five trials.

entation of the information. For *S*s in Group II the information was given 10 sec. after *S* completed the line, for *S*s in Group III 20 sec. after *S* completed the line, and for *S*s in Group IV 30 sec. after *S* completed the line. The *S*'s hand was returned to the starting point for the next trial after a time interval necessary to make the total time of 30 sec. for a trial. For *S*s in the control group no information was given and *S*'s hand was returned to the starting point 30 sec. after completing the last response.

RESULTS

The means and *SD*'s of the total number of "right" responses of each group in the 50 trials is presented in Table 1. A simple analysis of variance (1) showed that the effect of treatments was highly significant ($F = 58.75$, $df = 4$ and 35, $P = .001$). The mean differences were analyzed by t tests (1) for unmatched groups. The only difference between groups that was not significant at the .05 level or better was that between the 20-sec. and 30-sec. delay groups.

TABLE 1

MEANS AND *SD*'S OF "RIGHT" RESPONSES IN 50 TRIALS

	Delay in Seconds				Control
	0	10	20	30	
Mean	27.38	21.88	15.00	14.00	4.25
SD	4.12	4.26	2.00	2.06	1.48

The mean number of correct responses of each group for successive blocks of five trials is presented in Fig. 1. The consistency of the differences between groups throughout practice is clear.

DISCUSSION

The results of this experiment demonstrate the efficacy of immediate knowledge of results in the learning of a motor task. The results are in agreement with

the results obtained from the delay of reward in both human verbal learning and animal learning situations.

Though *S*s were instructed to maintain their hand at the point where they finished drawing the line, it is reasonable to assume that activity continued during the period before *S*'s hand was returned to the starting point. This activity would be concerned with the maintenance of the hand at the finishing point. This activity may differ from the activity of drawing the line. Hence the information which was delayed would be following the "hand-maintaining activity" rather than the "line-drawing activity." It is possible, therefore, that the information presented by *E* would tend to reinforce activity which may not be related to the "line-drawing activity" and thus reduce the effectiveness of the information. The one consistent activity was the "line-drawing activity." Thus even in the case of the 20-sec. and 30-sec. delay periods, there would be some activity associated with the "line-drawing activity" reinforced, but the incidence of stimuli eliciting the "right" response which were reinforced would be smaller with the increased delay and thus reduced the rate of learning. This position may be strengthened by the results of a similar experiment by Kanfer, Saltzman, and Greenspoon (3) in which significant differences were not obtained with increasing intervals of delay. In the latter experiment, however, the hand of *S* was not returned to the same starting point for successive trials. This may have enabled stimuli which were not present on the previous trial to be present on the following trial and thus affect the rate of learning regardless of the length of delay. Additional research, however, is necessary to resolve some of these issues.

Summary

The effect of delaying reward on the learning of a motor task by human *S*s was investigated. The task required *S*s to draw a 3-in. line while blindfolded. The information concerning the accuracy of the line was delayed for 0, 10, 20, and 30 sec. in four different experimental groups. No information was given to *S*s in the control group. The results indicate that increasing the length of the delay interval reduced the rate of learning. A delay up to 30 sec. was found to be superior to no information.

References

1. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
3. Kanfer, F. H., Saltzman, I. J., & Greenspoon, J. Delay of reward in motor learning. *Psychol. Rep.*, in press.
4. Perin, C. T. The effect of delayed reinforcement upon the differentiation of bar responses in white rats. *J. exp. Psychol.*, 1943, **32**, 95–109.
5. Perkins, C. C. The relation of secondary reward to gradients of reinforcement. *J. exp. Psychol.*, 1947, **37**, 377–392.
6. Saltzman, I. J. Delay of reward and human verbal learning. *J. exp. Psychol.*, 1951, **41**, 437–439.

(Received April 2, 1955)

Of Interest to Experimental Psychologists—

Contemporary Psychology

The *New* APA Journal of Book Reviews

The April issue of *CP* will feature a critical, essay-type review of *Stochastic Models for Learning* by Robert R. Bush and Frederick Mosteller. This detailed review by W. K. Estes, B. F. Skinner, and E. B. Wilson is presented from the different viewpoints of these three reviewers, one of whom comments that "It will be of more than academic interest to observe how students and teachers of learning respond to this friendly challenge from an adventuring physicist and a mathematical statistician." (An earlier issue of *CP* devoted nine pages to a review of *Modern Learning Theory*.)

- Every month *CP* presents many reviews of the important, new books in the psychological spectrum. And, these are all critical, evaluative reviews prepared by specialists in their particular fields. Let *CP* help you "keep up" with the new literature.

- Other *CP* features include: The Film Department—here's valuable information on all phases of educational films; "*CP* Speaks"—the editor's page of news and comment on developments in psychology's wide world of books; and "On the Other Hand"—a section reserved for comments by *CP*'s readers.

Contemporary Psychology belongs on your reading shelf. The subscription rate is $4.00 to APA members and participants of Student Journal Group; $8.00 to all others (add $.50 extra for foreign postage).

Send In Your Subscription *Now**
While Copies of Early Issues of *CP* Are Still Available

American Psychological Association

1333 Sixteenth Street N. W. **Washington 6, D. C.**

* If you wish, a letter *addressed to the attention of Dept. E31*, indicating your interest in a subscription to *CP* and mentioning your specific subscription group, *i.e.*, "APA member," will be sufficient. We'll then forward your subscription bill to you.